ABSTRACT

Title of Dissertation: ALGORITHM DEVELOPMENT FOR

                                HYPERSPECTRAL ANOMALY DETECTION.

                                Dalton S. Rosario, Doctor of Philosophy, 2008

Dissertation Directed By: Professor Ramalingam Chellappa,
                                Department of Applied Mathematics and
                                Scientific Computation and
                                Department of Electrical Engineering

        This dissertation proposes and evaluates a novel anomaly detection algorithm

suite for ground-to-ground, or air-to-ground, applications requiring automatic target

detection using hyperspectral (HS) data. Targets are manmade objects in natural

background clutter under unknown illumination and atmospheric conditions. The use

of statistical models herein is purely for motivation of particular formulas for

calculating *anomaly output surfaces*. In particular, formulas from semiparametrics are

utilized to obtain novel forms for output surfaces, and alternative scoring algorithms

are proposed to calculate output surfaces that are comparable to those of

semiparametrics. Evaluation uses both simulated data and real HS data from a joint

data collection effort between the Army Research Laboratory and the Army

Armament Research Development & Engineering Center.

        A data transformation method is presented for use by the two-sample data

structure univariate semiparametric and nonparametric scoring algorithms, such that,

the two-sample data are mapped from their original multivariate space to an univariate domain, where the statistical power of the univariate scoring algorithms is shown to be improved relative to existing multivariate scoring algorithms testing the same two-sample data.

An exhaustive simulation experimental study is conducted to assess the performance of different HS anomaly detection techniques, where the null and alternative hypotheses are completely specified, including all parameters, using multivariate normal and mixtures of multivariate normal distributions.

Finally, for ground-to-ground anomaly detection applications, where the unknown scales of targets add to the problem complexity, a novel global anomaly detection algorithm suite is introduced, featuring autonomous partial random sampling (PRS) of the data cube. The PRS method is proposed to automatically sample the unknown background clutter in the test HS imagery, and by repeating multiple times this process, one can achieve a desirably low cumulative probability of taking target samples by chance and using them as background samples. This probability is modeled by the binomial distribution family, where the only target related parameter—the proportion of target pixels potentially covering the imagery— is shown to be robust. PRS requires a suitable scoring algorithm to compare samples, although applying PRS with the new two-step univariate detectors is shown to outperform existing multivariate detectors.

ALGORITHM DEVELOPMENT FOR
HYPERSPECTRAL ANOMALY DETECTION.

By

Dalton S. Rosario

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Ramalingam Chellappa, Chair
Professor John J. Benedetto
Professor Eric V. Slud
Associate Professor Frank B. Alt
Associate Professor Adrian Papamarcou
Dr. Nasser M. Nasrabadi

Dedication

To the love of my wife Claudia, the remarkable spirit of adventure of our daughter

Pamela, and the analytical mind of our son Dalton II.

Acknowledgements

Table of Contents

List of Tables

List of Figures

Chapter 1   Introduction

1.1 Background

The field of spectroscopy examines the electromagnetic radiation that each unique material reflects, absorbs, and emits. The electromagnetic spectrum is sampled at a sufficiently large number of spectral bands to create a discrete spectral signature for different materials. In theory, each spectral signature should be unique for each unique material owing to its molecular structure. The ability to identify, within certain limits, physical materials from their spectral signature is the basis behind remote sensing imaging spectroscopy [1].

Remote sensing imaging spectroscopy involves using an airborne or space-borne platform with a sensor that records the reflected or emitted electromagnetic radiation. The sensor collects the radiation over a wide range of contiguous spectral bands, with each band corresponding to a unique spectral value. As the sensor moves above a region it records the electromagnetic radiation from a narrow swath of land, in many different spectral channels. The field of view of the sensor is broken into hundreds of thousands of pixels, with each pixel representing from less than one to many squared meters of the region of interest depending on the spatial resolution of the sensor and the height of the sensor during the data collection. A collection of spatial-spectral images is put together resulting in a hyperspectral (HS) data cube, where the length and width represent the spatial dimension, and the depth represents the spectral dimension [2].

The resulting HS data cube consists of hundreds of thousands of pixels. Each pixel has tens or hundreds of data points, each point corresponding to a unique spectral value. In theory, the spectral signature of each pixel should uniquely characterize the physical material in that spatial land area. In practice, the recorded spectral signatures will never be identical for samples of the same material. Owing to the different illumination conditions, atmospheric effects, sensor noise, etc., the resulting spectral signatures for HS data pixels containing similar materials will exhibit spectral variability.

1.2 Application of Statistical Models

Each spectral signature can be represented by a multidimensional vector, where each vector dimension represents a different spectral band. The spectrum of each pixel in a HS data cube is a vector lying in a multidimensional space. All pixels containing the same material and roughly the same amount of illumination will have their vector spectra closely grouped within the vector space, forming a sort of data cloud in the multidimensional space. The overall data space may contain many different homogeneous data clouds corresponding to the different materials in the HS data cube. Provided there are enough pixels in the data cube, or sample, this spectral variability can be modeled as a multivariate probability distribution.

Accurate models for the spectral variability of HS data are useful in many applications. Indeed, these models can be used to develop and evaluate algorithms for classification and detection, to select proper threshold, and to generate synthetic data.

Classification algorithms use probability distribution models to group the pixels of like material into spectrally homogeneous data sets. Classification algorithms label

each pixel in such a way that similar material pixels have the same label and the data can be segmented into spectrally homogeneous clusters. The material of each cluster can then be determined using available ground truth or comparing the statistical nature of the clustered pixels to that of a library of known materials.

Classification of HS image data has many applications. The classification labels can be used to determine the number of pixels of a particular material type in a scene, and since each pixel covers a spatial region, it is related to the amount of a material present. In commercial applications, this could be used, for instance, to compute farming yields, where the number of pixels of a specific crop could be used to estimate the amount of crop to be produced.

Detection algorithms use the probability distribution models to find pixels which contain a specific material of interest (target). The target pixels are considered rare relative to the number of pixels, which do not contain the target. Otherwise classification algorithms are used to segment the pixels. When the exact target signature is not known a priori, then the probabilistic models can be used to find pixels which are spectrally anomalous. In applications of anomaly detection algorithms, one tries to find objects that are significantly different spectrally from the other pixels within a scene.

Probabilistic models are also useful to generate synthetic HS data. The resulting synthetic data would have the same spectral variability as real world data and can be used to evaluate classification and detection algorithm under different conditions.

Classification and detection applications require accurate statistical models of the HS data to be effective. Without accurate models, algorithms performance is significantly reduced.

1.3 Statistical Models for Hyperspectral Data

Statistical signal processing uses a finite number of samples to model the probability of the data. The multivariate model defines the probability density function of the data. The effectiveness of a data model depends upon how accurately it represents all aspects of the data and how widely it applies.

Most utilized models are parametric, where the shape of a particular model is controlled by a set of parameters. If all of the parameters are known a priori, then the data model is completely known. When the parameters are not known a priori, they need to be estimated from the available data set.

For HS data, where the model and model parameters are not known a priori, typically a model for the spectral variability of the pixels is proposed and the model parameters are estimated using an entire HS dataset. The goal is to model the multivariate spectral variability of a particular data set as accurately as possible, where each pixel $\mathbf{x}_{rc} \in \mathbf{R}^{K} \left(1 \le r \le R; 1 \le c \le C \right)$ has K spectral bands, and the complete data cube $\mathbf{X} \in \mathbf{R}^{R \times C \times K}$ consists of RC pixels. The objective of statistical modeling of HS data is to propose a spectral probability distribution model and estimate its parameters

Given Multivariate Data:
$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11}, & \mathbf{x}_{12}, & \cdots, & \mathbf{x}_{1C} \\ \mathbf{x}_{21}, & \mathbf{x}_{22}, & \cdots, & \mathbf{x}_{2C} \\ & \vdots & & \\ \mathbf{x}_{R1}, & \mathbf{x}_{R2}, & \cdots, & \mathbf{x}_{RC} \end{bmatrix}$$

Proposed Multivariate Model: $\mathbf{x}_{rc} \sim f_K(\mathbf{x}; \Theta)$

Estimate Model Parameters: $\hat{\Theta}$

Since it is not possible to know the exact distribution of the data, real world processing of HS data must rely on limited information. The accuracy of the model parameter estimates depends on the number of HS data vectors. Due to the variation in spectral illumination, the different types of material present during each data collection, and other factors, only data collected during a single collection can be used to estimate the parameters of the multivariate model. In other words, the models must be determined in an adaptive manner from the HS data directly, with each HS dataset having different model parameters [3].

The primary model used for the spectral variability of HS image data is the multivariate normal distribution. While this model might do an adequate job modeling the main body of the data, rarely does it do an adequate job in modeling the tails of the distributions [3][4]. For detection applications at low probability of false alarms, with hundreds of thousands pixels in a HS data cube, to classify incorrectly a few hundred might not have a large effect on the overall classification of a scene. However, if, for example, a detection threshold is set with the expectation of one or two false alarms in the scene, and hundreds of pixels score higher than the threshold and are incorrectly labeled as targets, then the results of the detection algorithm become unreliable and of little value.

1.4 Relevant Work

In practice, it has been observed [5] that finding accurate models for high dimensional HS data may be unrealistic. Implementation of a parametric classifier for HS data is often cumbersome, it requires an unreasonably large set of training data to adequately characterize the multidimensional probability surface of each scene or target set, and it is difficult to store the description of such a surface unless it is well behaved [5]. It has also been observed [6] that it is extremely difficult to obtain an accurate density estimate non-parametrically in high-dimensional spaces. For these reasons, general class-conditional distribution functions are often replaced by a more tractable class-conditional distribution function for classification or detection using HS data.

Examples of parametric multivariate-normal-based target detection algorithms are the matched filter [7], the kernel spectral matched filter [8], and linear mixture models [9], [10], and [11]. Another main limitation of these representative parametric algorithms, in addition to assuming normality for HS data, is that they require a known target signature, and reliable target signatures are difficult to ascertain due to spectral variations already discussed in Section 1.1.

An alternative approach that does not require a spectral library for targets and has potential for invariance to atmospheric and illumination effects is anomaly detection, global or local.

Existing global anomaly detectors require that the HS data cube is first segmented into its constituent material classes. Then detection is achieved by applying a cutoff threshold and automatically locating pixel clusters with pixel values above the

threshold, representing the outliers of these classes. These hybrid algorithms vary in the method of segmentation, but tend to use maximum likelihood detection under the multivariate normal distribution. Furthermore, since the correct number of material classes in the scene is unknown a priori and needed by segmentation algorithms, this number is an unknown parameter that significantly affects the output results of such algorithms. The stochastic expectation maximization clustering algorithm [12] is a related example, see also [13].

Existing local anomaly detectors process small $(n \times n)$ windows of the HS data, where data sampling is not done in $\mathbf{X}$ (see Subsection 1.3); all the $\mathbf{x}_{rc}$ $(r = 1, \cdots, R; c = 1, \cdots, C)$ in $\mathbf{X}$ are used; modeling is only done at the level of the $n \times n$ windows, where $n \ll R$ and $n \ll C$ ($\ll$ denoting *many orders of magnitude smaller than*); and at the level of the pixel area surrounding these windows. Blocks of data ($n \times n$ windows) that are spectrally different from pixels surrounding them score high using an effective detector in contrast to blocks of data that are not spectrally different from their surrounding pixels. After the detector scores the entire $\mathbf{X}$, it yields a 2 dimensional (2-dim) surface Z [a $(R - n - 1) \times (C - n - 1)$ array of scalars], where a cutoff threshold is then compared to the pixel values in Z. Pixels having values greater than the threshold are labeled local anomalies. These are all features of existing anomaly detectors.

The most popular local anomaly detector in the HS research community is based on maximum likelihood estimation under the multivariate normal distribution; this detector is commonly known as the RX algorithm [14]. A *kernelized* version of RX has been also proposed [15]. For nonparametric local anomaly detection, the most

7

prominent multivariate detectors use classic methods, such as, Fisher's linear discriminant [15] and principal component decomposition [16][17].

Because local anomaly detectors (parametric or nonparametric) process small windows across the spatial area of $\mathbf{X}$, these algorithms are vulnerable to transitions across distinct regions in $\mathbf{X}$. Region transition events occur once a block of data representing a specific material is compared to a surrounding mixture of pixels representing the same material and one or more additional, but distinct, material types. This sort of events can augment the probability of false alarms in $\mathbf{X}$ because a spectral sample consisting of pixels of two or more material types is, indeed, different from a spectral sample consisting of pixels of a single material. Since existing local anomaly detectors do not directly account for local transitions of distinct regions, there is a need to address the problem.

Finally, local anomaly detectors are limited to applications where the scales of targets in $\mathbf{X}$ (relative sizes of targets to other objects in the imagery) are expected to be known a priori. This prior knowledge is available in air to ground (top view) detection applications, where the sensors look straight down at the ground at a known altitude. However, this prior knowledge is not available in ground-to-ground (ground view) detection applications, where target scales are dependent on the range between sensors and targets. If the goal is to detect targets as spectral local anomalies in the scene, one has to ensure that a small window in the imagery (inside window) is reasonably separated from its surrounding pixel region (outside window) to avoid having a block of target data compared to surrounding pixels that also belong to the same target. So, using the inside-outside window method for sample comparison,

properly setting the separation gap between the inside and outside windows must be done a priori and is a critical factor, completely removing ground view anomaly detection applications, as candidate applications using inside-outside windows. An alternative sampling method is needed for ground view anomaly detection applications.

HS image data offer clear advantages over conventional broadband images—each pixel has K bands in HS image data versus one band in broadband images, but with current detection algorithm vulnerabilities, the topic of robust target detection is still open for research.

1.5 Overview of This Work

This dissertation focuses on the development and evaluation of algorithm suites for ground view (GV) and top view (TV) anomaly detection applications using HS data cubes. In this context, a target is any manmade object in a natural clutter background, whose spectral signature is not available, and if available, is considered unreliable and is not used in the approach.

An algorithm suite consists of a host of techniques each performing a specific task in order to achieve the overall goal of detecting, autonomously, the presence of targets in the scene as spectral anomalies in the HS imagery, yielding in the process a low false alarm probability.

If targets are present in the scene, each target is assumed to be represented by multiple pixels $\mathbf{x}_{rc}$ ($\mathbf{x}_{rc} \in \mathbf{R}^{K}; 1 \leq r \leq R; 1 \leq c \leq C$, see Subsection 1.3) in data cube $\mathbf{X}$ ($\mathbf{X} \in \mathbf{R}^{R \times C \times K}$), and cover an area in $\mathbf{X}$ greater than or equal to $n \times n$, where $n << R$ and $n << C$.

This dissertation analyzes $\mathbf{X}$ for TV anomaly detection applications by sliding a $n \times n$ inside window and testing the observed spectral sample against surrounding spectra (outside window), as described in Subsection 1.4 for local anomaly detectors. For GV anomaly detection applications, this dissertation also analyzes $\mathbf{X}$ by sliding a $n \times n$ window, but proposes to test the observed spectral sample in the $n \times n$ window against N randomly selected $n \times n$ blocks of data taken from $\mathbf{X}$. The latter testing approach addresses the uncertainty on target scales, as discussed in Subsection 1.4 for local anomaly detectors, by eliminating the need for an outside window. It also automatically addresses the global anomaly detection problem without the need to use unreliable segmentation techniques in $\mathbf{X}$, as described in Subsection 1.4 for global anomaly detectors.

However, as also discussed in Subsection 1.4, any testing approach that uses sliding windows is vulnerable to transitions across distinct spectral regions in $\mathbf{X}$.

This dissertation establishes that using a data transformation method that maps multivariate spectral samples to univariate samples, and applying univariate detectors to test the transformed samples can significantly reduce the probability of false alarms in $\mathbf{X}$ compared to multivariate anomaly detectors.

In order to show a fair comparison between the multivariate anomaly detection techniques and the two-step univariate anomaly detection techniques (i.e., data transformation step followed by a univariate scoring step), a real HS data cube having a target satisfying the assumptions stated in this section is first chosen and characterized using human aided segmentation to establish useful $n \times n$ window (spatial) models, and later used for assignment of probabilistic models (under the

10

multivariate normal distribution family) for the only data-structure used by the detectors in this work—a two-sample data structure. The parametric spectral models and null and alternative hypotheses are formulated at the level of the $n \times n$ window models and used for simulation studies, where both multivariate and univariate anomaly detection approaches are evaluated using a standard statistical method to estimate the power of correct detection and the type II error, given a type I error and sample size fixed to n^2.

Since the only data-structure used by the detectors in this work is a two-sample data structure, detectors are compared using simulated two-sample data (generated independently of each other) that are based on multivariate normal distributions and mixtures of multivariate normal distributions. Parameters are estimated using the segmented image and the real HS image data for the multivariate normal distributions of the different spectral groups in the segmented image. During each trial in the simulation experiments, the two-sample data $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n^2} \in \mathbf{R}^K$ and $\left\{ \mathbf{y}_{2u} \right\}_{u=1}^{n^2} \in \mathbf{R}^K$ — representing the spectral sample observed via the sliding window and a reference spectral sample—are independently generated and shared by all multivariate anomaly detectors (older detectors) chosen for these experiments. The two-sample data are also transformed to two univariate samples $\left\{ x_{1h} \right\}_{h=1}^{n^2} \in \mathbf{R}$ and $\left\{ x_{2u} \right\}_{u=1}^{n^2} \in \mathbf{R}$, where a data transformation method (Chapter 3 shows details) is introduced to address the need for invariance to the illumination environment and certain atmospheric conditions (e.g., lower visibility, fog). Both univariate samples $\left\{ x_{1h} \right\}_{h=1}^{n^2}$ and $\left\{ x_{2u} \right\}_{u=1}^{n^2}$

are shared by all univariate anomaly detectors (new detectors) proposed in this dissertation.

Most of the older anomaly detectors described in this work enters in the comparison analysis only as applied to the *K*-dim HS data, while the new detectors operate only on the transformed data.

The use of statistical models for the development of detectors described in this work is purely for motivation of particular formulas for calculating *anomaly output surfaces*. In particular, formulas from semiparametrics are utilized to obtain novel forms for output surfaces, and alternative scoring algorithms are also proposed to calculate output surfaces that are comparable to those of semiparametrics, using the same HS dataset.

Finally, this dissertation presents a fully operational GV anomaly detection algorithm suite and evaluates the suite using real HS data cubes, where targets are present in a natural clutter background under different illumination and atmospheric conditions. The data were recorded during a recent joint data collection effort between Army Research Laboratory and Army Armament Research, Development and Engineering Center.

1.6 Significance of This Work

To date, a significant amount of research has focused on classification and detection algorithms using HS image data, while little has been done to address the underlying fundamental problems that affect algorithm performances, and act on them. It is beyond the scope of this dissertation to address all of the underlying fundamental problems in classification and detection algorithms, as highlighted in

Subsection 1.4, but the ones addressed directly in this work for HS anomaly detection are important steps in the right direction, as discussed in Subsection 1.5 and shown later on in this dissertation.

This dissertation directly yields the following contributions:

- Introduction of a novel GV global anomaly detection algorithm suite, featuring autonomous partial random sampling of the data cube. The random sampling method is modeled by the binomial distribution family. Parametric or nonparametric segmentation is no longer necessary to achieve effective global anomaly detection.

- Introduction of a HS data transformation method for sliding window based tests that maps multivariate samples to univariate samples, reducing the effects of the illumination and atmospheric conditions on the discriminant power among spectral samples of distinct material types.

- The first use of formulas from semiparametrics on HS image data in order to obtain novel forms for output surfaces; also, alternative scoring algorithms are proposed to calculate output surfaces that are comparable to those of semiparametrics, using the same HS data set.

- The first use of simulated null and alternative hypothesis tests to assess sliding window based HS anomaly detection algorithms. Simulation experiments for estimation of the type I error and power of correct detection are conducted in twofold: (i) using the data structure of two idealized sample data and (ii) idealized multispectral data cubes, where samples are based on a multivariate normal distribution or mixtures of multivariate normal distributions.

1.7 Organization of Dissertation

The remainder of this dissertation is organized as follows:

Chapter 2 covers various relevant materials for later chapters, and fixes the notations for most of the discussions through out those chapters. The SOC-700 HS imager is described, the sensor used to record the HS data tested in Chapter 7. A brief description on the physics of HS sensing is presented making the reader familiar with the physical measurement recorded by this sensor modality. A real HS data cube is segmented and characterized in this chapter by making assumptions and estimating parameters that will be used for simulation experiments in later chapters.

Chapter 3 proposes a procedure that takes calibrated spectral samples of different sizes $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$ and $\{\mathbf{y}_{2u}\}_{u=1}^{n_2} \in \mathbf{R}^K$, in units of radiance per band, and maps them to $\{x_{1h}\}_{h=1}^{n_2} \in \mathbf{R}$ and $\{x_{2u}\}_{u=1}^{n_2} \in \mathbf{R}$, where $0^0 \leq \{x_{1u}\}_{u=1}^{n_2} \leq 90^0$ and $0^0 \leq \{x_{2u}\}_{u=1}^{n_2} \leq 90^0$ (both having the same size—n_2) in units of angular degree. This chapter presents experimental results, using the Kolmogorov-Smirnov test on transformed data, that assesses whether transformed blocks of data are independent and identically distributed when compared to randomly selected spectra (after transformation) from a large data set representing a single material type.

Chapter 4 presents the adaptation of multiple multivariate and univariate techniques to the anomaly detection problem using HS data. First, it presents a brief discussion on multivariate normal based statistical hypothesis testing, describes the state of the art in HS anomaly detection prior to this work, and presents alternative multivariate anomaly detection techniques. This chapter adapts univariate techniques

14

for the first time to HS anomaly detection applications, which includes the application of a semiparametric model and other alternative data combining metrics.

Chapter 5 describes and quantifies the effects of spectral magnitude (bias), spectral shape, and spectral mixtures (heterogeneous samples) have on the power performances of multivariate and univariate anomaly detection techniques. The experiment outcome are put in context to what is desired and undesired in anomaly detection applications—this discussion should shed some lights on later performances of multivariate and univariate detectors on real HS imagery.

Chapter 6 demonstrates under varying TV background configuration scenarios the differences of performances between multivariate and univariate anomaly detectors using idealized data cubes. Differently from the simulation experiments discussed in Chapter 5, the simulation experiments discussed in this chapter generate idealized five-band data cubes, and test for local anomalies using the sliding inside-outside dual window sampling method. Note that component proportions in mixtures observed by the windows do not need to be specified, as in Chapter 5, because they would occur naturally as the inside-outside dual window slides across the spatial areas of the artificial data.

Chapter 7 introduces a parallel (repeated) random sampling approach and models this approach using the Binomial distribution family. It discusses how this sampling approach can be implemented in the context of GV anomaly detection, and presents results using real GV HS imagery.

Chapter 8 summarizes this work, offers limitations and some thoughts on future work.

Chapter 2   HS Sensing, Data Characterization and Models

2.1 Background

This chapter covers various relevant materials for later chapters in this dissertation, including fixing the notations for most of the discussions through out those chapters. Section 2.2 describes a brief history on the evolution of HS sensors, including some of the details on well known HS sensors currently being deployed. The SOC-700 HS sensor is of particular interest, because data from this sensor were used for experiments results that will be discussed later herein. Section 2.3 presents a brief description on the physics of HS sensing so that the reader can better appreciate the data recorded by this sensor modality. Section 2.4 characterizes a real HS data cube by making assumptions and estimating parameters that will be used later on for simulation analysis. Section 2.5 summarizes this chapter.

2.2 Hyperspectral Sensors

In the past 35 years, the field of imaging spectrometry has undergone tremendous development. Imaging spectrometry refers to the imaging of a scene over a large number of discrete, contiguous spectral bands in order to obtain a complete reflectance spectrum from the imaged ground surface. This type of imaging is also known as hyperspectral imaging. In 1972, NASA launched the first experimental Earth Resources Technology Satellite (ERTS). It was the first in a line of remote sensing satellites that have been renamed the Landsat series. The first satellites carried an instrument called the Multi-Spectral Scanner (MSS). It provided repeated coverage of the Earth through an inclined polar orbit that captured images of Earth in

4 broad bands of the electromagnetic spectrum. The final image produced by ground processing of the instrument-data typically yielded a picture very similar to that produced by false-color infrared film (also known as camouflage-detection film). MSS satellite images provide synoptic views of the earth that don't require meticulous mosaic-ing of many individual scenes. However, each digital picture element (pixel) only resolves an object about 80 meters (250 feet) in diameter.

A decade later, an instrument that offered both higher spatial and spectral resolution – called the Thematic Mapper (TM) – was launched on the Landsat 4 satellite. The TM instrument resolves objects down to less than 30 meters (100 feet) and also adds the capability of imaging the ground in 3 additional broad bands in the infrared portion of the electromagnetic spectrum; this includes one low-resolution thermal band.

With the advent of TM and more sophisticated computer software, it is possible to gain more insight into the minerals and hence the types of rocks present. Not only can ferrous oxide powder be recognized, but also clays derived from alteration can be identified more reliably than before, by comparing bands (ratioing). Vegetation can be discriminated from other surface features by comparing bands of high and low reflectance. Additional regions of interest for ground-verification can be identified through what is called multispectral classification. Ground-truth or photo-interpretation is used to identify representative land classes and the computer is instructed to search for regions that appear similar in all bands.

In 1986, SPOT Image (a French company) launched its first commercial remote sensing satellite. Its most notable contribution to remote sensing is panchromatic

(B&W) images with a resolution of 10 meters (33 feet) per pixel. It also provides multispectral imagery, similar to the Landsat MSS instrument, with 20 meter (66 feet) resolution.

We discuss next, three specific HS systems that are well known in the HS research community: AIS, AVIRIS, and SOC-700.

Hyperspectral Systems

Airborne Imaging Spectrometer (AIS):

The Airborne Imaging System (AIS 1, AIS 2) was designed and built in the early 1980's as part of a NASA Jet Propulsion Laboratory (JPL) imaging spectrometry program [18]. This instrument was designed explicitly for multispectral infrared imaging and used 32 x 32 element mercury cadmium telluride area detector array with 10-bit quantization. AIS used a grating spectrometer with push-broom style scan to separate the signal into 128 contiguous bands in the spectral region from 1.2 to 2.4 μm with spectral resolution of 9.3 nm. The spectra were sampled sufficiently fine for analysts to identify spectral features of specific minerals for unambiguous classification. The early success of AIS enabled NASA to upgrade the instrument (AIS II) with a 64 x 64 element HgCdTe array that extends the spectral range up to the visible range covering 0.8 to 2.4 μm, but the performance was limited by 7.3° FOV, low spatial resolution, and the fact that it was not radiometrically calibrated.

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS):

In late 1980's, the next generation of imaging spectrometer in the NASA program was the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Flown

first in 1987 and in current use, AVIRIS is designed to image 224 contiguous bands in the spectral region from 0.4 μm to 2.5 μm with about 0.01 μm resolution. The increased spectral range in the visible region and high resolution compared to AIS enable it to detect important absorption features in vegetation and minerals such as the shifts in the chlorophyll and the kaolinite doublet at 2.2 μm. AVIRIS is designed to have an altitude of about 20 km for an GIFOV of 20 m over a swath width of 12 km [19]. In order to cover relatively wide spectral range and narrow spectral bandwidth, AVIRIS uses four spectrometers, one with a silicon array for the visible and near IR region and the other with InSb array for SWIR region [19].

Imaging spectrometer data from AVIRIS have been applied to many other uses in the field of atmospheric science, botany, hydrology, oceanography and remote sensing. The main focus of these applications has been in the identification, measurement, and monitoring of constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures.

Surface Optics Corp 700 (SOC-700) Hyperspectral Sensor:

More recently, better signal to noise ratio, higher spectral resolution, and lower cost HS systems have been introduced to the market and are commercially available. The SOC-700 HS imager [20] is one of those systems, and it was the system that recorded all the real HS image data used for the work in this dissertation, see Fig. 2.1 and examples of real HS data cubes in Chapter 7.

The SOC-700 system comprises of a high-speed, low-noise visible camera; a high quality visible spectrometer; integrated scanning system; and vector processor software. This system can record and process HS imagery at a rate of 64,000 120-

band elements per second at 12-bit resolution. The system's spectral response covers

the visible and near infrared (VNIR) spectral range from 0.43 to 0.97 μm and can be

used in normal to low lighting conditions with variable exposure times. The system

can be configured to operate either as an imager or line scanner, producing up to 640

pixels per line, 640 lines. Under favorable conditions, the SOC-700 camera has a line

rate of up to 100 lines per seconds (120 wavelength bands). With six programmable

correlation channels, three integrated channels, on-the-fly dark frame subtraction and

calibration, the system can be configured to display color, panchromatic, simulated

sensor and detect or reject with a relatively high degree of confidence up to 6

spectrally identified elements.

Figure 2.1. SOC-700 Hyperspectral Imaging System.

The SOC-700 provides the following output dataset to the user: raw data cubes,

calibrated data cubes, correlation cubes, RGB or sensor simulated cubes. This

systemcan be operated on a local machine or remotely over a local area network.

Communication between the GUI (graphics user interface) and the system are accomplished through sockets and multiple cameras can be controlled from the same machine.

The heart of the SOC-700 system is the so-called MP-1 Midis Processor which can perform on-the-fly hyperspectral processing at super computer speeds. Capable of processing 128,000 128-band vector per second, the MP-1 can match the processing requirements of any currently available hyperspectral imaging system and can be configured to work with most systems.

2.3 Hyperspectral Sensing Model

This subsection briefly describes a simplified model for the HS reflectance phenomenology. This model, although not explicitly used in this dissertation, does provide a connection between a spectral sample, which is treated herein as an ordinary multivariate sample, and the physics behind the information recorded by a HS device.

As discussed earlier, HS data are produced by a sensor that either scans or uses a focal plane array to collect the data in a rectangular grid about the region of interest. The sensor filters the data in such a manner as to provide a large number of narrow wavelength bands. Each pixel then represents a resolution spot size on the ground. In order to appreciate how the atmospheric and illumination conditions affect the reflectance of an object in the ground, consider a relationship derived in [21] for the spectral radiance reaching an airborne or satellite sensor, it can be expressed in simplified form as

$$L_b = \int_\Gamma B_b \left[\left(GE_s(\lambda)\tau_1(\lambda)\cos\theta + FE_d(\lambda) \right)\tau_2(\lambda)\frac{\delta(\lambda)}{\pi} + L_u(\lambda) \right] d\lambda \qquad (2.1)$$

where Γ is the spectral region of interest centered at λ_b (the central wavelength in the

b^{th} band in units of μm), L_b is the effective spectral radiance in the bth band in units of

$[Wcm^{-2}s\,\delta^{-1}\mu m^{-1}]$, $E_s(\lambda_b)$ is the exoatmospheric spectral irradiance from the Sun in

units of $[Wcm^{-2}\mu m^{-1}]$, $\tau_1(\lambda_b)$ is the transmission through the atmosphere along the

Sun-object path, θ is the angle from the surface normal to the Sun, F is the fraction of

the spectral irradiance from the sky $[E_d(\lambda_b)]$ incident on the object (i.e., not blocked

by adjacent objects), G is the fraction of direct sunlight incident on the object, $\tau_2(\lambda_b)$

is the transmission along the object-sensor path, $\delta(\lambda_b)$ is the spectral reflectance

factor for the object of interest (i.e., $\delta(\lambda_b)/\pi$ is the bidirectional reflectance in units of

$s\,\delta^{-1}$), $L_u(\lambda_b)$ is the spectral path radiance $[Wcm^{-2}s\,\delta^{-1}\mu m^{-1}]$, and β_b is the normalized

spectral response of the bth spectral band of the sensor under study where

$$B_b = \frac{\rho_b(\lambda_b)}{\int_\Gamma \rho_b(\lambda)d\lambda} \qquad (2.2)$$

with $\rho_b(\lambda_b)$ being the peak normalized spectral response in Γ of the bth band.

Atmospheric and illumination conditions will affect all the radiometric terms in (2.1)

(i.e., $E_s(\lambda_b)$, $\tau_1(\lambda_b)$, $\tau_2(\lambda_b)$, $E_d(\lambda_b)$, and $L_u(\lambda_b)$), which makes the task of predicting the

responses of a particular object a formidable one. For a particular set of conditions

during the data collection, the spectral radiance from a pixel-size location at the scene

observed by a K-band sensor can be expressed as

$$\mathbf{x} = [L_1, L_2, \cdots, L_b, \cdots, L_K], \qquad (2.3)$$

where scalars L_k $(k = 1, \cdots, K)$ are radiances, such that, adjacent radiances (e.g., L_b and L_{b+1}) are usually highly correlated.

Using (2.3), a HS data cube $\mathbf{X} \in \mathbf{R}^{R \times C \times K}$, a $R \times C$ array of K bands, is constructed as

$$
\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11}, & \mathbf{x}_{12}, & \cdots, & \mathbf{x}_{1C} \\ \mathbf{x}_{21}, & \mathbf{x}_{22}, & \cdots, & \mathbf{x}_{2C} \\ & & \vdots & \\ \mathbf{x}_{R1}, & \mathbf{x}_{R2}, & \cdots, & \mathbf{x}_{RC} \end{bmatrix}, \tag{2.4}
$$

where, $\mathbf{x}_{rc} \in \mathbf{R}^K$ is an observation vector at pixel row r $(r = 1, \cdots, R)$ and column c $(c = 1, \cdots, C)$.

2.4 Data Characterization

This section uses a real HS data cube of the form in (2.4) and characterizes small blocks of data across the spatial area $R \times C$ of $\mathbf{X}$. It is worth noting that data sampling is not done in $\mathbf{X}$, all the $\mathbf{x}_{rc}$ in $\mathbf{X}$ are used, and that modeling is only done at the level of small windows (blocks of data). This is relevant to anomaly detection algorithms because they observe data using small windows across $\mathbf{X}$. The data cube used for this section was collected by the SOC-700 sensor (Section 2.2), and the scene is considered a difficult one for anomaly detection applications. Data characterization is covered herein in three subsections: Subsection 2.4.1 (observation probability estimation in $\mathbf{X}$), Subsection 2.4.2 (spatial window modeling), Subsection 2.4.3 (null/alternative hypothesis modeling), and Subsection 2.4.4 (parameter specification). These results will be used later on in Chapter 5 to conduct simulation experiments.

2.4.1 Event Probabilities Using Small Windows in X

As modeling is done at the level of small windows in $\mathbf{X}$, certain events, such as, the observation of samples of two or more different material types, or samples of the same material under different illumination conditions (e.g., shaded and non-shaded tree regions), can play a major role degrading performance of anomaly detectors. These events contradict a popular assumption made for scoring metrics that a block of data consists only of a single material type under the same illumination condition. In practice, this assumption is not always satisfied using real HS imagery, and, because it has been ignored in the HS research community, anomaly detectors have been known to yield a high false alarm rate; hence, their utility has been limited to preliminary imagery screening tasks.

Figure 2.2 shows the 2-dim display version of a real HS data cube having the format of $\mathbf{X}$ (2.4), where *R = C = 640* and *K = 120*.

The scene depicted in Fig. 2.2 was recorded using the SOC-700 VNIR HS sensor from a ground to ground viewing perspective. The scene consists of an open grassy field, trees, bushes, and a motor vehicle in tree shades—a sport utility vehicle (SUV), which is encircled in Fig 2.2. From the available ground truth information, this SUV (the target) had three of its doors open, including a side-opening backdoor. Additional details about this scene from the available ground truth information are given in Chapter 7.

Figure 2.2. Ground to ground HS scene.

The image shown in Fig. 2.2 represents the radiance sample average per pixel location, or

$$\overline{\mathbf{X}} = \begin{bmatrix} \overline{L}_{11}, \ \overline{L}_{12}, \ \cdots, \ \overline{L}_{1C} \\ \overline{L}_{21}, \ \overline{L}_{22}, \cdots, \ \overline{L}_{2C} \\ \vdots \\ \overline{L}_{R1}, \ \overline{L}_{R2}, \cdots, \ \overline{L}_{RC} \end{bmatrix} \tag{2.5}$$

where, $\overline{L}_{rc} = \dfrac{1}{K} \sum\limits_{k=1}^{K} L_{rck}$ and L_{rck} is the radiance on the kth frequency band at pixel row

$r \ (1 \leq r \leq R)$ and column $c \ (1 \leq c \leq C)$.

If we use an $n \times n$ window, such that $n \ll R$ and $n \ll C$ ($\ll$ denoting *much smaller than*), to define *events* within $n \times n$ blocks of data in (2.5), then the events can be defined in terms of the pixel values within the whole $n \times n$ window (through a rule that 5% or more of the pixels in a window must be in a group for the group to

25

have *occurred* within that window), and the event is then ascribed to a pixel that is the *location* (the upper-left corner) of that window.

For instance, letting *ij* index the left upper corner pixel in this window relative to the indexation shown in (2.5), we could set an index to automatically count the number *m* of *ij* locations in the image a particular event occur within the $n \times n$ window, where a particular event could be the presence of spectral samples belonging to two different material classes (e.g., tree leaves and a particular paint of a motor vehicle), as *i* and *j* step through $1, \cdots, (R - n - 1)$ and $1, \cdots, (C - n - 1)$, respectively, across the image. Using *m*, the probability of this event's occurrence within the $R \times C$ image, using the $n \times n$ window, is then the relative frequency of pixels within a whole data cube of falling into a designated category, or $\dfrac{m}{(R - n - 1)(C - n - 1)}$.

The goal of this subsection is to measure such probabilities for a number of different events, which will be described shortly.

Let m_e be the number of the eth event $(e = 1, 2, \cdots, E)$ that can occur across an $R \times C$ image within a fixed $n \times n$ window. We are interested in computing the event probabilities f_e (relative frequencies), as follows:

$$f_e = \frac{m_e}{(R - n - 1)(C - n - 1)}. \tag{2.6}$$

To simplify the computation of (2.6), we must first segment the image shown in Fig 2.2. A note about *segmentation*: since the problem of image segmentation is not a solved problem, segmentation techniques discussed in the literature (see, for instance,

[22]) are highly sensitive to parameters set by the user (e.g., the desired number of different classes in the imagery), we used a combination of an automatic segmentation technique and manual editing in order to separate four distinct spectral *groups* in Fig 2.2. A group is defined for this purpose as regions in the image consisting of spectra that are relatively closer to one of the few most distinct spectral means in the imagery. (Note: single pixels do not belong to more than one group.) In particular, since most of the anomaly detectors use spectral mean averages to detect anomalies in the imagery, we used the standard *k-means* approach [22] to decompose the scene into multiple spectrally distinct groups. Since the number of groups is a parameter using the k-means approach, this parameter was changed from 7 to 3, but the target could not be isolated as a group. So, the number of group parameter was set to 3 and manual editing was required to isolate the target as a fourth group. In addition, since a material class under different illumination condition can play an important role on performance degradation of anomaly detectors, samples of a material class under similar illumination condition may be considered as a different event relative to samples of the same material under a different illumination condition.

Fig. 2.3 shows the final 640×640 segmented image, which has the following format:

$$\mathbf{A} = \begin{bmatrix} a_{11}, & a_{12}, & \cdots & , & a_{1C} \\[6pt] a_{21}, & \overbrace{\begin{bmatrix} a_{ij}, & a_{i(j+1)}, & \cdots, a_{i(j+n-1)} \\ a_{(i+1)j}, & a_{(i+1)(j+1)}, & \cdots, a_{(i+1)(j+n-1)} \\ & \vdots & \\ a_{(i+n-1)j}, & a_{(i+n-1)(j+1)}, & \cdots, a_{(i+n-1)(j+n-1)} \end{bmatrix}}^{n \times n \ \text{window, where in this case } i=j=2} & \begin{matrix} ,\cdots, & a_{2C} \\ ,\cdots, & a_{(2+1)C} \\ \vdots & \vdots \\ ,\cdots, & a_{(2+n-1)C} \end{matrix} \\ a_{(2+1)1}, & & \\ \vdots & & \\ a_{(2+n-1)1}, & & \\ \vdots & \vdots & \\ a_{R1}, & a_{R2}, & \cdots & , & a_{RC} \end{bmatrix} . \quad (2.7)$$

where $a_{rc} \in \{1, 2, 3, 4\}$ is the observed group number at pixel row r $(r = 1, \cdots, R)$ and column c $(c = 1, \cdots, C)$, and the block area in $\mathbf{A}$ labeled with "$n \times n$ window, where in this case $i = j = 2$," has the pixel at this window's left upper corner indexed by ij, where $i = 1, 2, \cdots, (R - n - 1)$ and $j = 1, 2, \cdots, (C - n - 1)$. (The $n \times n$ window is shown in (2.7) because event probability estimations require the use of such a window in order to record the observed event at location ij.)

Groups 1 through 3 (G1, G2, and G3) represent natural clutter background regions in the imagery, and group 4 (G4) represents the target.

With four distinct groups, we define seven events for observations viewed through a window: E1, E2, E3, E4, E5, E6, and E7; where E1 denotes samples of a single group (G1, G2, G3, or G4); E2 denotes a mixture consisting of samples from any two groups (e.g., G1 and G3); E3 denotes a mixture of any three groups; E4 denotes a mixture of the four groups; E5 denotes any combination of 2 or 3 groups in a mixture that *excludes* G4 (the target); E6 denotes any combination of 2 or 3 groups

in a mixture that *includes* G4; and E7 denotes observation of G4 only. (A group to be counted as observed must occupy at least 5% of the spatial window.)

Table 2.1 shows the estimated probabilities of these events using (2.6) and a 20×20 window at *ij* locations across the segmented image, where $i = 1,2,3,\cdots,(R-19)$, $j = 1,2,3,\cdots,(C-19)$, and $R = C = 640$. The window area corresponds to 0.0977% of the $R \times C$ segmented image.
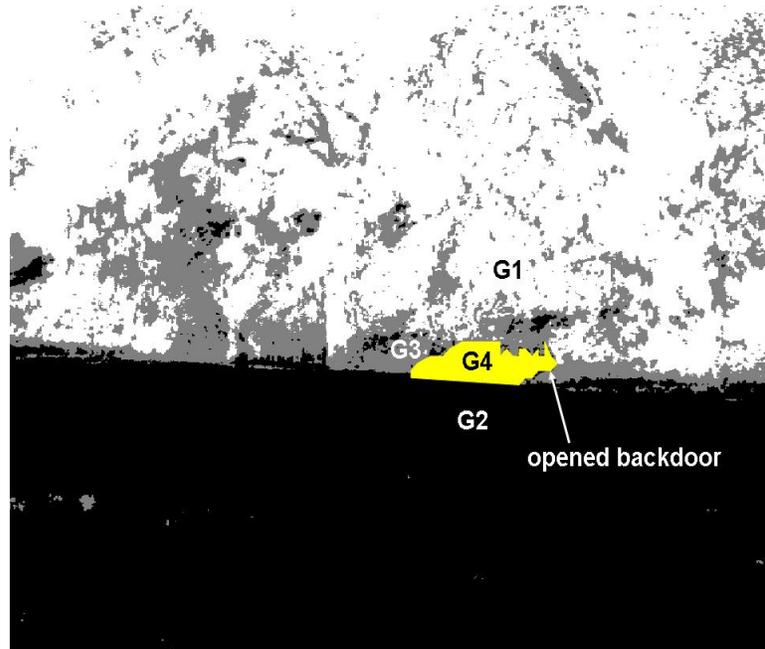
Figure 2.3. Segmented HS data cube using the *k-means* approach and manual editing to isolate the target as a fourth group (G4). Notice that resulting segmentation using pixels of 120 bands may not correspond to results by human inspection of Fig. 2.2. Groups G1, G2, and G3 correspond to natural clutter background.

These events will be used in Subsection 2.4.2 for window spatial modeling, and in Subsection 2.4.3 for null and alternative hypothesis modeling.

The cumulative event probabilities in Table 2.1 involving E2, E3, and E4, although relatively small when compared to the probability of event E1, can play a major negative role in the value of using anomaly detectors, as these detectors assume that E1 models all overlapping blocks of data in $\mathbf{X}$. As discussed in Section 1.3, $\mathbf{X}$ is expected to have hundreds of thousands of pixels.

Table 2.1. Event Probabilities Using a Segmented HS Data Cube.

Observed Event	Event Probability
E1 (single group)	0.76883078322104
E2 (mixture of 2)	0.13975357487286
E3 (mixture of 3)	0.08565095594334
E4 (mixture of 4)	0.00576468596276
E1+E2+E3+E4	1.00000000000000
E5 (mixture of 2 or 3, excluding G4)	0.08664897098450
E6 (mixture of 2 or 3, including G4)	0.01957448300362
E7 (G4)	0.01407498476562

In the example shown in Table 2.1, there seems to be an accepted truth classifying each $n \times n$ block of data as belonging to one or more groups of background clutter (G1 to G3) and/or target group (G4). We define mutually exclusive labels E1 to E4 by defining pixels according to their categories of group membership (G1 to G4). But

there is implicitly a finer subdivision, which we can make use of in defining E5 to E7, according to whether G4 is or is not one of the groups a pixel belongs to. Thus, event probability is simply the relative frequency of pixels within a whole data cube of falling into a designated category.

The most important fact about the results shown in Table 2.1 is that group mixtures (E2, E3, E4, E5, and E6) are always present in real imagery, and, as it will be shown in Chapter 5 and Chapter 6 (simulation experiments) and in Chapter 7 (experiments using real HS imagery), this presence degrades performance of the most commonly used multivariate anomaly detectors—false positives increase.

The specific application will dictate the level of tolerance for false positives. An application requiring, for instance, that an image analyst take decisions upon viewing the output of anomaly detectors asks for an extremely low number of false positives, because false positives are often found scattered across the imagery forming isolated pixel clusters of target scales. In many applications, targets occupy less than 0.1% of the imagery, while blocks of data consisting of mixtures may occupy a larger portion of the imagery relative to targets.

The ability to deal effectively with sample mixtures is one of the goals in this work. Events E5 (potential to increase false alarms) and E6 (potential to decrease power of correct detection) are of particular interest. We will model E5 to train anomaly detectors and use data modeled for E6 for testing. In practice, anomaly detectors are not required to be trained, but for analysis purposes, training will be used.

2.4.2 Data Models for Small Windows

This subsection presents models for observed data using a sliding $n \times n$ window in $\mathbf{X}$. These models will be used in Subsection 2.4.3, where further determinations will be made to establish null and alternative hypotheses.

The data format of $\mathbf{X}$ is shown in (2.4), where r $(r = 1, \cdots, R)$ and c $(c = 1, \cdots, C)$ index pixels $\mathbf{x}_{rc}$ in the $R \times C$ spatial area $\mathbf{X}$, where $n << R$ and $n << C$. Pixels within a fixed $n \times n$ block of data in $\mathbf{X}$ (i.e., data observed through a $n \times n$ window) are indexed from the upper left corner of this block using ij relative to rows and columns in $\mathbf{X}$, where $i = 1, \cdots, (R - n - 1)$ and $j = 1, \cdots, (C - n - 1)$. A representation of an $n \times n$ window at pixel location $(i, j) = (2,2)$ in $\mathbf{X}$ is

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11}, & \mathbf{X}_{12}, & & \cdots & & , & \mathbf{X}_{1C} \\ & & \overbrace{\qquad n \times n \ \text{window} \ , \text{where} \ \ \text{in this} \ \ \text{case} \ \ i = j = 2 \qquad} & & & & \\ \mathbf{X}_{21}, & \left[\mathbf{X}_{ij}, \right. & \mathbf{X}_{i(j+1)}, & \cdots & \left. , \mathbf{X}_{i(j+n-1)} \right], & \cdots, & \mathbf{X}_{2C} \\ \mathbf{X}_{(2+1)1}, & \mathbf{X}_{(i+1)j}, & \mathbf{X}_{(i+1)(j+1)}, & \cdots & , \mathbf{X}_{(i+1)(j+n-1)}, & \cdots, & \mathbf{X}_{(2+1)C} \\ \vdots & & & \vdots & & \vdots & \vdots \\ \mathbf{X}_{(2+n-1)1}, & \mathbf{X}_{(i+n-1)j}, & \mathbf{X}_{(i+n-1)(j+1)}, & \cdots, & \mathbf{X}_{(i+n-1)(j+n-1)}, & \cdots, & \mathbf{X}_{(2+n-1)C} \\ \vdots & & & \vdots & & & \vdots \\ \mathbf{X}_{R1}, & \mathbf{X}_{R2}, & & \cdots & & , & \mathbf{X}_{RC} \end{bmatrix} . \quad (2.8)$$

As the $n \times n$ window slides across $\mathbf{X}$ in (2.8), different spatial events are observed. Using the segmented image in Fig. 2.3 as a reference, an illustration of different spatial models of these events is shown in Fig 2.4.
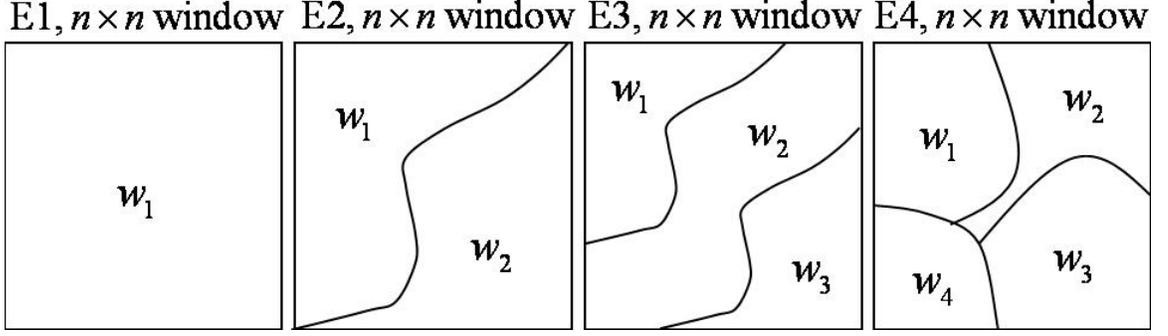
Figure 2.4 Spatial models for event observations viewed by an $n \times n$ window across $\mathbf{X}$, where $\left\{ w_a = \left(G1, G2, G3, G4 \right) \right\}_{a=1}^{4}$ and $w_1 \neq w_2 \neq w_3 \neq w_4$. Events E2, E3, and E4 correspond to mixtures of 2, 3, and 4 groups, respectively.

Before pixels within a block of data can be used by a detector, they need to be rearranged to a sequence of multivariate samples. The rearrangement is made by concatenating individual rows in the $n \times n$ window in (2.8), as follows

$$\mathbf{W}_1 = \left[\mathbf{x}_{ij}, \cdots, \ \mathbf{x}_{i(j+n-1)}, \ \mathbf{x}_{(i+1)j}, \cdots, \ \mathbf{x}_{(i+1)(j+n-1)}, \cdots, \mathbf{x}_{(i+n-1)j}, \ \cdots, \ \mathbf{x}_{(i+n-1)(j+n-1)} \right]$$

$$= \left[\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right],$$

$$(2.9)$$

where $\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}$, $n_1 = n^2$, and $\mathbf{y}_{1h} \in \mathbf{R}^K \left(h = 1, \cdots, n_1 \right)$, such that

$\mathbf{y}_{11} = \mathbf{x}_{ij}$, $\mathbf{y}_{12} = \mathbf{x}_{i(j+1)}$ and so forth until finally $\mathbf{y}_{1n_1} = \mathbf{x}_{(i+n-1)(j+n-1)}$. Since a window can be anywhere in $\mathbf{X}$ and $\mathbf{X}$ represents any HS data cube, $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n_1}$ are considered random vectors.

From the dimensions of the window and spatial area of $\mathbf{X}$, $i = 1, \cdots, \left(R - n - 1 \right)$ and $j = 1, \cdots, \left(C - n - 1 \right)$ in $\mathbf{X}$. So, as the location of the left upper corner pixel in the

observation window changes by an unit within $\mathbf{X}$, i.e., $\mathbf{y}_{11} = \left\{\mathbf{x}_{ij}\right\}_{i=1;\,j=1}^{R-n-1;\,C-n-1}$, the entire

set of spectra that constitutes $\mathbf{X}$ will be observed through the $n \times n$ window.

Using the assumption that random vectors in $\mathbf{W}_1$ are independent and

identically distributed (i.i.d.), and using the data cube in Fig. 2.2 as reference, the

distribution of data within the window, using (2.9), can be simplified to

$$\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \sim \text{i.i.d. } g_1(\mathbf{y} \mid \boldsymbol{\theta}) = \sum_{a=1}^{4} g(\mathbf{y} \mid \boldsymbol{\theta}_a)\rho_a \tag{2.10}$$

where $g(\mathbf{y} \mid \boldsymbol{\theta}_a)$ is a group-conditional PDF with unknown parameter set $\boldsymbol{\theta}_a$,

$a \in \{1,2,3,4\}$ indexes the spectral group, and ρ_a is the unknown proportion of

$g(\mathbf{y} \mid \boldsymbol{\theta}_a)$ contributing to $g_1(\mathbf{y} \mid \boldsymbol{\theta})$, such that $0 \le \rho_a \le 1$ and $\sum_{a=1}^{4} \rho_a = 1$, and $\boldsymbol{\theta}$ is the

parameter set $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4)$.

We further simplify (2.10) by letting $g(\mathbf{y} \mid \boldsymbol{\theta}_a)$ be a family of normal PDFs,

$N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$, such that, both parameters—mean $\boldsymbol{\mu}_a \in \mathbf{R}^K$ and covariance $\boldsymbol{\Sigma}_a \in \mathbf{R}^{K \times K}$—

are unknown.

Groups G1, G2, G3, and G4, discussed in Subsection 2.4.1, are distributed

under $g(\mathbf{y} \mid \boldsymbol{\theta}_1)$, $g(\mathbf{y} \mid \boldsymbol{\theta}_2)$, $g(\mathbf{y} \mid \boldsymbol{\theta}_3)$, and $g(\mathbf{y} \mid \boldsymbol{\theta}_4)$, respectively. And, by a well

known property of the normal distribution, $g_1(\mathbf{y} \mid \boldsymbol{\theta})$ in (2.10) is also normal, since it

is the result of a sum of normal PDFs. Notice also that $\{\rho_a\}_{a=1}^{4}$ in (2.10) determines

whether $\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1}$ is a multivariate sample belonging to a single group, e.g.,

$g_1(\mathbf{y} \mid \boldsymbol{\theta}) = g(\mathbf{y} \mid \boldsymbol{\theta}_3)$, or belonging to a mixture of groups, e.g.,

$g_1(\mathbf{y} \mid \boldsymbol{\theta}) = g(\mathbf{y} \mid \boldsymbol{\theta}_1)\rho_1 + g(\mathbf{y} \mid \boldsymbol{\theta}_4)\rho_4$, where in this example $\rho_2 = \rho_3 = 0$ and

$\rho_1 + \rho_4 = 1$. Spatial examples of mixtures are illustrated in Fig. 2.4.

The data model shown in (2.10) will be used to establish null and alternative hypotheses for later use in Chapter 5 for simulation experiments.

2.4.3   Data Models for Null and Alternative Hypotheses

In order to decompose subtle factors that affect performances of different anomaly detection techniques, flexible hypothesis tests (null and alternative) must be modeled and specified for simulation experiments. This subsection addresses this topic.

In defining a hypothesis test, we must first consider that an anomaly detector requires two input sets of spectra $\left(\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}\right)$ and $\left(\mathbf{W}_2 \in \mathbf{R}^{K \times n_2}\right)$ to perform its task using $\mathbf{X}$. Independently of the application, or viewing perspective between sensor and scene, one of the two inputs $\left(\mathbf{W}_1\right)$ is obtained at a fixed location *ij* in $\mathbf{X}$, as shown in (2.8) and (2.9); but the other input set will depend on the application, or viewing perspective. For instance, for the ground viewing perspective shown in Fig. 2.2, the second input set $\mathbf{W}_2$ could be made available from a spectral library, or be randomly selected straight out of the testing image cube. In either case, $\mathbf{W}_2$ would be a rearranged version of a $n \times n$ block of data. We address both cases in detail in Chapter 7, where, in order to make such a test useful for real applications, we propose that $\mathbf{W}_1$ be independently compared to multiple spectral sets,

$\mathbf{W}_2^{(f)} \in \mathbf{R}^{K \times n_2}$ $\left(f = 1, \cdots, N\right)$, and results from this comparison fused, as it will be shown in Chapter 7.

Another sampling method is to use pixel vectors surrounding a $n \times n$ block of data to construct $\mathbf{W}_2$, where $\mathbf{W}_1$ is constructed from the block of data. Both input sets $\mathbf{W}_1$ and $\mathbf{W}_2$ feed the anomaly detector. Although this latter approach is suitable for top viewing perspectives, it is unsuitable for ground viewing perspectives, as discussed in Section 1.4. (Chapter 6 describes in detail implementation of this inside-outside window method.)

Whether the perspective is ground view or top view, mixtures of different groups in $\mathbf{W}_1$ and/or in $\mathbf{W}_2$ can significantly degrade anomaly detectors' performances, as discussed in Section 1.4. But since the mixture problem has been ignored by the HS research community, we could not find guidance in the literature on modeling the problem using a standard statistical method for permutation tests of significance.

After several considerations, including the fact that target group G4 is in tree shades, we settled for modeling the mixture problem using combinations of E5 (mixture of 3, excluding G4) and E1 (G3 only)—see Table 2.1—to represent a difficult H_0, and combinations of E5 (mixture of 3, excluding G4) and E6—see Table 2.1—to represent H_1. We define next the null and alternative hypotheses, where calibration of the rejection thresholds for all anomaly detectors considered in this work, for a desired type I error, are discussed in Chapter 5. In particular for the hypotheses, using (2.10) as reference, the following null and alternative hypotheses will be considered as baselines for simulation experiments:

$$\text{NULL } H_0 \quad \begin{aligned} \mathbf{W}_2 &= \left(\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2} \right) \ \sim i.i.d. \ \ g_2(\mathbf{y} \mid \boldsymbol{\theta}) = \sum_{a=1}^{3} g(\mathbf{y} \mid \boldsymbol{\theta}_a) \pi_a \\ \mathbf{W}_1 &= \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right) \ \sim \ i.i.d. \ \ g_1(\mathbf{y} \mid \boldsymbol{\theta}) = g(\mathbf{y} \mid \boldsymbol{\theta}_3) \end{aligned} \quad (2.11)$$

where, $\left\{ g(\mathbf{y} \mid \boldsymbol{\theta}_a) \right\}_{a=1}^{3}$ are assumed multivariate normal PDF's with parameters given in

Table 2.2 [see also comments on (2.10)], $0 \le \pi_a \le 1$ and $\sum_{a=1}^{3} \pi_a = 1.0$, $\left\{ \mathbf{y}_{2u} \right\}_{u=1}^{n_2} \in \mathbf{R}^K$

are observation vectors available through a spectra library (or randomly selected)

representing event E5 (G1, G2, and G3), $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n_1} \in \mathbf{R}^K$ are observation vectors at

pixels in an $n \times n$ window (each pixel with K spectral bands), such that $\mathbf{y}_{11} = \mathbf{x}_{ij}$ (the

left upper corner pixel of the window) and ij index rows and columns in $\mathbf{X}$ such that

only data from G1 can be observed [using notation in (2.10),

$\rho_1 = 1, \rho_2 = 0, \rho_3 = 0,$ and $\rho_4 = 0$], and $\mathbf{W}_1$ is assumed independent of $\mathbf{W}_2$.

Parameter specification for $\left\{ g(\mathbf{y} \mid \boldsymbol{\theta}_a) \right\}_{a=1}^{3}$ will be discussed in Subsection 2.4.4.

In essence, using (2.11) as the null hypothesis will enable us to study the

vulnerability of anomaly detectors through simulation experiments using the most

difficult null hypothesis that Fig. 2.2 can offer, i.e., compare spectra from a single

group $g(\mathbf{y} \mid \boldsymbol{\theta}_3)$ to spectra from a mixture of three groups $\sum_{a=1}^{3} g(\mathbf{y} \mid \boldsymbol{\theta}_a) \pi_a$, such that,

$g(\mathbf{y} \mid \boldsymbol{\theta}_3)$ is included in the mixture.

For the H_0 in (2.11), it is desired that anomaly detectors can keep their cutoff

thresholds at a low value relative to the corresponding output results of these

detectors on the following alternative hypothesis:

ALTERNATIVE H_1
$$\mathbf{W}_2 = (\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2}) \ \sim i.i.d. \ \ g_2(\mathbf{y}|\boldsymbol{\theta}) = \sum_{a=1}^{3} g(\mathbf{y}|\boldsymbol{\theta}_a)\pi_a$$
$$\mathbf{W}_1 = (\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1}) \ \sim \ i.i.d. \ \ g_1(\mathbf{y}|\boldsymbol{\theta}) = \sum_{a=1}^{4} g(\mathbf{y}|\boldsymbol{\theta}_a)\rho_a \qquad (2.12)$$

where $\mathbf{W}_2$ are rearranged pixels from the surrounding area of the $n \times n$ window or

from a spectral library or from a randomly selected block of data in $\mathbf{X}$, $\mathbf{W}_1$ are

rearranged pixels from the $n \times n$ window representing E6 or E7 (depending on group

proportions), group proportions $\{\rho_a\}_{a=1}^{4}$ sum to one, $i = 1, \cdots, (R-n-1)$ and

$j = 1, \cdots, (C-n-1)$; and $\mathbf{W}_1$ is assumed independent of $\mathbf{W}_2$. Parameter specification

for the normal distributions $\sum_{a=1}^{3} g(\mathbf{y}|\boldsymbol{\theta}_a)\pi_a$ and $\sum_{a=1}^{4} g(\mathbf{y}|\boldsymbol{\theta}_a)\rho_a$ will be discussed in

Subsection 2.4.4.

It is worth noting that varying proportions $\{\rho_a\}_{a=1}^{4}$ in (2.12) significantly

changes the difficulty level of H_1 relative to H_0, as it will be shown in Chapter 5.

Parameter specification will be discussed next.

2.4.4   Parameter Specifications

This section specifies parameters for the group conditional models shown in

(2.11) and (2.12), which will be used for simulation experiments in Chapter 5.

Models (2.11) and (2.12) show three groups representing natural clutter background

$\{g(\mathbf{y}|\boldsymbol{\theta}_a)\}_{a=1}^{3}$ and a forth group representing a target $g(\mathbf{y}|\boldsymbol{\theta}_4)$. Since $g(\mathbf{y}|\boldsymbol{\theta}_a)$

$a \in \{1,2,3,4\}$ is a group-conditional normal PDF having unknown parameter set $\boldsymbol{\theta}_a$,

$\boldsymbol{\theta}_a$ were approximated by $(\hat{\boldsymbol{\mu}}_a, \hat{\boldsymbol{\Sigma}}_a)$—sample mean $\hat{\boldsymbol{\mu}}_a \in \mathbf{R}^{120}$ and sample covariance

$\boldsymbol{\Sigma}_a \in \mathbf{R}^{120 \times 120}$—using the segmented image shown in Fig 2.3 to mask out the

corresponding group regions in $\mathbf{X}$. The HS data cube depicted in Fig. 2.2 was $\mathbf{X}$ for these estimations. The target group (G4) used about 6,000 pixel vectors for parameter estimations, and each one of the other groups used a significantly higher number of pixel vectors for their parameter estimations, especially G1 and G2. But because of restrictions imposed by the Army Research Laboratory, we arbitrarily chose to use the sample covariance of G3 for both G3 and G4. Proportions $0 \le \pi_a \le 1$ and

$0 \le \rho_a \le 1$ will also be specified in this section, such that, $\sum_{a=1}^{3} \pi_a = 1$ and

$$\sum_{a=1}^{4} \rho_a = 1.$$

The estimated covariance matrices for G1, G2, and G3 are displayed in Fig. 2.5, so that one can visually appreciate the radiance correlation among the 120 frequency bands for each group.

Estimated means and estimated variances corresponding to these groups are partially shown in Table 2.2 (only the first 10 components are shown for illustration purposes).

Group contributions to a mixture, proportions (π_1, π_2, π_3) and $(\rho_1, \rho_2, \rho_3, \rho_4)$, will be specified based on the simulation purpose. Simulation experiments that will use H_0 in (2.11), for instance, will arbitrarily use fixed proportions $(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ in order to expose the detectors to a hard H_0 in $\mathbf{X}$, see Fig. 2.2. Based on our own experience working with HS data, an easier H_0 involving mixture would not have utility for real applications. On the other hand, using more than four groups in H_0 for $\mathbf{W}_2$ would most likely increase the difficulty level for the anomaly detectors discussed in this dissertation. So, instead of simulating more

variations than the ones presented in this dissertation, we chose to complement the

results produced by simulation experiments with results yielded on real HS data. The

latter is presented in Chapter 7.
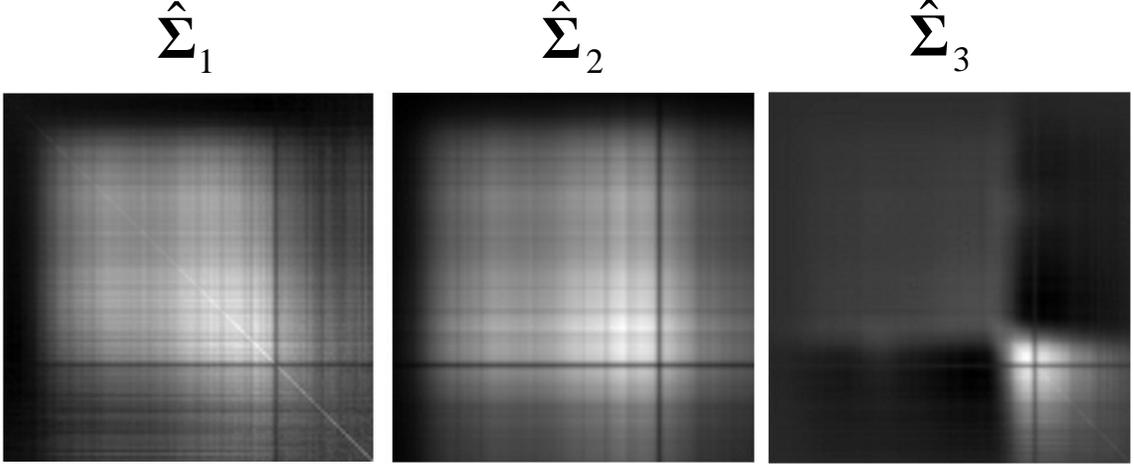
Figure 2.5. Estimated covariances for G1 $\left(\hat{\mathbf{\Sigma}}_1\right)$, G2 $\left(\hat{\mathbf{\Sigma}}_2\right)$, and G3 $\left(\hat{\mathbf{\Sigma}}_3\right)$ of dimensions 120×120 displayed as intensity images after linear mapping the gray scale of each to the range 0-255. The upper left-hand corner of each matrix is the estimated response variance at frequency band 1; the lower right-hand corner is the estimated variance at band 120. In simulation experiments, $\hat{\mathbf{\Sigma}}_4$ (G4) will be set to $\hat{\mathbf{\Sigma}}_3$, for reasons explained in the text.

For the alternative hypothesis H_1 in (2.12), proportion parameters

$\left(\rho_1, \rho_2, \rho_3, \rho_4\right)$ determine the level of difficulty imposed on the detector by H_1 with

respect to H_0. For instance, H_1 would correspond to H_0 by setting

$\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(1.0, 0.0, 0.0, 0.0\right)$, and by using small variations on this setting, e.g.,

$\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(0.9, 0.0, 0.0, 0.1\right)$, for $\hat{\mathbf{\mu}}_4 \neq \hat{\mathbf{\mu}}_k$ $\left(k = 1, \cdots, 3\right)$, H_1 would represent a

hard alternative hypothesis for a detector to detect the target; alternatively, by setting

$\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(0.0, 0.0, 0.0, 1.0\right)$, for $\hat{\mathbf{\mu}}_4 \neq \hat{\mathbf{\mu}}_k$ $\left(k = 1, \cdots, 3\right)$, H_1 would represent an

easy alternative hypothesis using the same detector. In particular, we discuss in Chapter 5 simulation experiment results setting

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.02, 0.02, 0.01, 0.95),$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.04, 0.03, 0.03, 0.90),$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.07, 0.07, 0.06, 0.80),$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.14, 0.13, 0.13, 0.60), \tag{2.13}$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.20, 0.20, 0.20, 0.40),$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.27, 0.27, 0.26, 0.20), \text{ and}$$

$$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.32, 0.32, 0.31, 0.05).$$

Justification for (2.13) and additional settings will be discussed in Chapter 5.

It is worth noting that $\hat{\boldsymbol{\mu}}_4$ is relatively close to $\hat{\boldsymbol{\mu}}_3$. This is because the target, as shown in Fig. 2.2, is under tree shades. We use this fact to study in detail the detection performance of various detectors as a function of varying magnitude or shape of $\hat{\boldsymbol{\mu}}_4$, everything else fixed. To perform this study, we will modify in Chapter 5 the target mean vector by applying two parameters to this mean: one that affects only the overall magnitude (or bias) of $\hat{\boldsymbol{\mu}}_4$, and another that affects only the shape of $\hat{\boldsymbol{\mu}}_4$ (see Chapter 5).

Table 2.2. (First 10 Bands Only) Estimated means $\hat{\boldsymbol{\mu}}_1$ (G1), $\hat{\boldsymbol{\mu}}_2$ (G2), $\hat{\boldsymbol{\mu}}_3$ (G3), and $\hat{\boldsymbol{\mu}}_4$ (G4). Estimates $diag(\hat{\boldsymbol{\Sigma}}_1)$, $diag(\hat{\boldsymbol{\Sigma}}_2)$, and $diag(\hat{\boldsymbol{\Sigma}}_3)$ are the diagonal terms of $\hat{\boldsymbol{\Sigma}}_1$, $\hat{\boldsymbol{\Sigma}}_2$, and $\hat{\boldsymbol{\Sigma}}_3$, respectively. Dimensions are 120×1.

$\hat{\boldsymbol{\mu}}_1$ (1.0e3)	$\hat{\boldsymbol{\mu}}_2$ (1.0e3)	$\hat{\boldsymbol{\mu}}_3$ (1.0e3)	$\hat{\boldsymbol{\mu}}_4$ (1.0e3)	$diag(\hat{\boldsymbol{\Sigma}}_1)$ (1.0e3)	$diag(\hat{\boldsymbol{\Sigma}}_2)$ (1.0e5)	$diag(\hat{\boldsymbol{\Sigma}}_3)$ (1.0e4)
0.204325	0.175925	0.087980	0.079486	0.160109	0.004161	0.011063
0.220555	0.190895	0.094050	0.092472	0.121826	0.005398	0.010110
0.244355	0.216750	0.102410	0.108234	0.138863	0.007945	0.012112
0.260325	0.236395	0.106035	0.118299	0.158180	0.011452	0.014576
0.282800	0.260630	0.111530	0.105730	0.172251	0.016150	0.015648
0.303000	0.285110	0.114955	0.118079	0.192190	0.022174	0.018835
0.306015	0.295890	0.116005	0.109305	0.155532	0.028275	0.018890
0.343840	0.329070	0.128410	0.115210	0.195692	0.039375	0.025303
0.400310	0.385030	0.145650	0.166185	0.229260	0.063204	0.033793
0.426200	0.414335	0.152510	0.174521	0.253849	0.080857	0.039194

2.5 Summary

This chapter fixed many of the notations for later chapters in this dissertation. For the sliding window testing approach, which is the only one used in this dissertation, all observation vectors in $\mathbf{X}$ are used, and models and null and alternative hypotheses were formulated at the level of the small windows in $\mathbf{X}$.

Definitions were given to distinguish a material spectral class from a spectral group, where a material spectral class refers to spectra of a single material type (e.g., asphalt, a particular color paint) under the same illumination condition, and a spectral group refers to regions in the imagery having *similar* spectra—independently of the material type (for instance, two material types under the same illumination condition may belong to the same group, and the same material type under different illumination conditions may belong to different groups). This distinction was necessary in order to isolate subtle factors that might degrade anomaly detection

performances. These factors will be analyzed in detail in Chapter 5. Ideally, spectral sets of the same material type under different illumination conditions should not be scored as anomalous to each other.

A challenging HS data cube $\mathbf{X}$ (i.e., target in tree shades) for anomaly detection applications was selected and segmented in four spectral groups, were only four groups were chosen to simplify the segmentation process. Event probabilities were computed using a 20×20 window across the segmented version of $\mathbf{X}$. Events were defined to distinguish different spatial observations using this window. Table 2.1 shows probability estimates of these events. The most important fact about these probabilities is that group mixtures (E2, E3, E4, E5, and E6) are expected to exist in real HS imagery. Chapter 5 and Chapter 6 will show detailed analyses on the effect these mixtures have on anomaly detection performances.

Small windows in $\mathbf{X}$ were then modeled as a sequence of i.i.d. random multivariate variables under the sum of four normal PDFs, where parameters were estimated using the segmented image to mask out the four groups in $\mathbf{X}$. Using this model, null and alternative hypotheses were specified. These hypotheses will be used later in later chapters for assessment of various anomaly detectors.

Chapter 3   HS Data Transformation

3.1 Introduction

This chapter starts by making a distinction between HS data preprocessing and HS data transformation.

Data preprocessing has a specific meaning in the HS research community, it may consist of (i) rectification of system and sensor distortions in data as it is received directly from the Space Segment in preparation for delivery (raw data); (ii) registration of such data with respect to features of Earth; and (iii) calibration of spectral response with respect to such data, but not manipulation or further calculation with such data, or combination such data with other data. Data preprocessing leads to calibrated data—a spectral data product produced by applying corrections to collected data so that sets in physically interpretable quantities result (e.g., flux, radiance, energy and temperature, which relate to the scene or object). The data model presented in (2.1), and the subsequent data cube in (2.4), is a simplified form of calibrated data, or preprocessed data.

In this dissertation, the goal of data transformation is to convert multivariate samples of preprocessed HS data to univariate samples in order to heighten contrasts between the known different types of terrain and targets.

The remainder of this chapter is organized as follows. Section 3.2 proposes a procedure that takes calibrated spectral samples of different sizes $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$ and $\{\mathbf{y}_{2u}\}_{u=1}^{n_2} \in \mathbf{R}^K$, in units of radiance per band, and map them to $\{x_{1h}\}_{h=1}^{n_2} \in \mathbf{R}$ and

$\{x_{2u}\}_{u=1}^{n_2} \in \mathbf{R}$, where $0^0 \leq \{x_{1u}\}_{u=1}^{n_2} \leq 90^0$ and $0^0 \leq \{x_{2u}\}_{u=1}^{n_2} \leq 90^0$ (both having the same

size—n_2) in units of angular degree. Section 3.3 presents an i.i.d. test experiment

using the Kolmogorov-Smirnov test [23] on transformed data, as the detection models

used in this dissertation assume i.i.d. sequences. Finally, Section 3.4 summarizes and

concludes the chapter.

3.2 A Data Transformation Method

We propose a data transformation approach in two steps: (i) spectral differencing

and (ii) angle mapping. The rationale for (i) is twofold: (*a*) since HS samples are

contiguous in the spectral domain (i.e., typical spectral resolution is of the order of 10

nanometers), we believe that more discriminant information can be found between

adjacent bands, which could augment the statistical power of detectors; and (*b*)

differencing spectra should significantly decrease the importance of spectral

magnitude (or bias) in anomaly detection applications, while significantly increasing

the importance of spectral shapes. Spectral magnitude relates to the mean average of

all measured radiance within a spectral sample, and spectral shape relates to the

plotted curve of measured radiance as a function of frequency band (Chapter 5 shows

details.) Existing classification and detection algorithms directly or indirectly exploit

magnitude and/or shape of spectra in order to perform their tasks.

Different from the prior art, however (see, for instance, [4]), this dissertation

treats these features separately and independently, thus, exposing their strengths and

weaknesses, see Chapter 5.) The benefit of (ii) is that it reduces the multivariate

problem to a univariate problem, which avoids the problem of singularity during

inverse estimations of covariance matrices. Singularity is known to occur (see, for

instance, [4]) when the sample size of $\mathbf{W}_2$ and/or $\mathbf{W}_1$ [see, for instance, (2.11) or (2.12)] is smaller than K (the number of bands) in the HS data cube [see (2.3)]. Depending on the application, some of the targets may be relatively small, consisting of only a few pixels (e.g., sample size of $\mathbf{W}_1$) , which would be relatively small compared to K. A typical HS sensor usually delivers between 120 and 1,000 bands, while targets may vary in number of pixels from as large as in the thousands to as small as *1* to *4* pixels, depending on the actual physical sizes of these targets and/or distance between the sensor and targets.

The two-step data transformation approach is discussed next.

Borrowing from the discussion in Section 2.4, this data transformation approach requires two sets of spectra, $\mathbf{W}_1$ and $\mathbf{W}_2$. (Recall that $\mathbf{W}_1$ is a multivariate sequence of spectra rearranged from a block of data in $\mathbf{X}$, and $\mathbf{W}_2$ is another multivariate sequence rearranged from a spectral library, from a randomly selected block of data in $\mathbf{X}$, or rearranged from the area surrounding the $n \times n$ window). Sequences $\mathbf{W}_1$ and $\mathbf{W}_2$ are represented in radiance form as

$$
\mathbf{W}_2 = \begin{bmatrix} \mathbf{y}_{21}, & \cdots, & \mathbf{y}_{2n_2} \end{bmatrix} \qquad \sim \; g_2(\mathbf{y} \mid \mathbf{\eta})
$$

$$
= \begin{bmatrix} L_{211}, & \cdots, L_{21n_2} \\ L_{221}, & \cdots, L_{22n_2} \\ L_{231}, & \cdots, L_{23n_2} \\ & \vdots \\ L_{2(K-2)1}, & \cdots, L_{2(K-2)n_2} \\ L_{2(K-1)1}, & \cdots, L_{2(K-1)n_2} \\ L_{2K1}, & \cdots, L_{2Kn_2} \end{bmatrix} \tag{3.1}
$$

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{y}_{11}, & \cdots, & \mathbf{y}_{1n_1} \end{bmatrix} \quad \sim \quad g_1(\mathbf{y}\,|\,\boldsymbol{\theta})$$

$$= \begin{bmatrix} L_{111}, & \cdots, & L_{11n_1} \\ L_{121}, & \cdots, & L_{12n_1} \\ L_{131}, & \cdots, & L_{13n_1} \\ & \vdots & \\ L_{1(K-2)1}, & \cdots, & L_{1(K-2)n_1} \\ L_{1(K-1)1}, & \cdots, & L_{1(K-1)n_1} \\ L_{1K1}, & \cdots, & L_{1Kn_1} \end{bmatrix}, \qquad (3.2)$$

where neither the forms of $g_2(\mathbf{y}\,|\,\boldsymbol{\eta})$ and $g_1(\mathbf{y}\,|\,\boldsymbol{\theta})$ (joint PDFs) nor their parameters $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ are known, and L_{ukh} are radiance values (scalars), as shown in (2.3).

The magnitude of L_{ukh} depends on the amount of illumination and the illumination environment. This dependence can be virtually eliminated by applying a first order difference—an approximation of the derivative—to the columns of $\mathbf{W}_1$ and $\mathbf{W}_2$, or

$$\nabla_2 = \begin{bmatrix} (L_{221} - L_{211}), & \cdots & , (L_{22n_2} - L_{21n_2}) \\ (L_{231} - L_{221}), & \cdots & , (L_{23n_2} - L_{22n_2}) \\ & \vdots & \\ (L_{2(K-1)1} - L_{2(K-2)1}), & \cdots, & (L_{2(K-1)n_2} - L_{2(K-2)n_2}) \\ (L_{2K1} - L_{2(K-1)1}), & \cdots, & (L_{2Kn_2} - L_{2(K-1)n_2}) \end{bmatrix} \qquad (3.3)$$

$$\nabla_1 = \begin{bmatrix} (L_{121} - L_{111}), & \cdots & ,(L_{12n_1} - L_{11n_1}) \\ (L_{131} - L_{121}), & \cdots & ,(L_{13n_1} - L_{12n_1}) \\ & \vdots & \\ (L_{1(K-1)1} - L_{1(K-2)1}), & \cdots, & (L_{1(K-1)n_1} - L_{1(K-2)n_1}) \\ (L_{1K1} - L_{1(K-1)1}), & \cdots, & (L_{1Kn_1} - L_{1(K-1)n_1}) \end{bmatrix}. \tag{3.4}$$

Notice in (3.3) and (3.4) that $\nabla_1 \in \mathbf{R}^{(K-1)\times n_1}$ and $\nabla_2 \in \mathbf{R}^{(K-1)\times n_2}$. The sample

averages of ∇_1 and ∇_2 are, respectively,

$$\overline{\nabla}_2 = \frac{1}{n_2} \nabla_2 \mathbf{1}_{n_2 \times 1} \tag{3.5}$$

and

$$\overline{\nabla}_1 = \frac{1}{n_1} \nabla_1 \mathbf{1}_{n_1 \times 1}, \tag{3.6}$$

where $\mathbf{1}_{d \times 1}$ is a column vector of dimension d.

If we denote $\mathbf{W}_2$ the reference sample and the columns of its corresponding

∇_2 [i.e., the columns of (3.3), which has $K-1$ dimensions] as $\left\{\nabla_{2u} \in \mathbf{R}^{(K-1)}\right\}_{u=1}^{n_2}$, then

multivariate samples can be mapped to univariate samples as

$$x_{2u} = \frac{180}{\pi} \arccos \left(\frac{\nabla_{2u}^t \overline{\nabla}_2}{\|\nabla_{2u}\| \|\overline{\Delta}_2\|} \right), \tag{3.7}$$

$$x_{1u} = \frac{180}{\pi} \arccos \left(\frac{\nabla_{2u}^t \overline{\nabla}_1}{\|\nabla_{2u}\| \|\overline{\nabla}_1\|} \right), \tag{3.8}$$

where $0^0 \le \{x_{1u}\}_{u=1}^{n_2} \le 90^0$, $0^0 \le \{x_{2u}\}_{u=1}^{n_2} \le 90^0$, the operator $\|\mathbf{x}\|$ using a column vector

$\mathbf{x}$ denotes the square root of $\mathbf{x}^t\mathbf{x}$.

From (3.7) and (3.8), two univariate sequences are constructed

$$x_2 = \left(x_{21},\ x_{22},\ \cdots,\ x_{2n_2}\right) \ \sim\ f_2(x\,|\,\eta) \tag{3.9}$$

and

$$x_1 = \left(x_{11},\ x_{12},\ \cdots,\ x_{1n_2}\right) \ \sim\ f_1(x\,|\,\theta), \tag{3.10}$$

where $f_2(x\,|\,\eta)$ and $f_1(x\,|\,\theta)$ are unknown joint PDFs having unknown parameter sets

η and θ, x_2 (reference) and x_1 (test) are used as input sequences for the univariate

based anomaly detection techniques that will be discussed in Chapter 4.

3.3 I.I.D. Test Experiment Using Transformed Data

Anomaly detection techniques based on the assumption of i.i.d. samples are

common in the target community. Our approach to anomaly detection also relies on

this assumption; but, differently from other approaches, we assume that transformed

calibrated data samples are i.i.d.—not the actual calibrated data samples. This section

aims at checking the plausibility of i.i.d. data-transformed samples from $n \times n$ blocks

of data in $\mathbf{X}$—in particular those of a single spectral group. In order to check this

plausibility, we will compare the normalized histogram of randomly selected spectra

of a single material type against the normalized histograms of spectra from blocks of data belonging to the same material type.

We will first segment real HS data cubes, and then select a spectral group that covers a large area in these cubes. Pixels of *K* bands representing this group will be randomly selected from a large population of pixels, and the cumulative collection will be rearranged into a single sequence as in (3.1), and finally transformed using (3.3), (3.5) and (3.7). The distribution of the resulting sequence, which has the structure of (3.9), will be empirically estimated. We will then select $n \times n$ blocks of data that are significantly apart from each other representing the same spectral group; pixels from each block will be rearranged as in (3.1), transformed using (3.3), (3.5), and (3.7), and also have their corresponding distributions empirically estimated. Estimated distributions using blocks of transformed pixels will be compared to the estimated distribution using transformed randomly selected pixels.

If the data transformation method produces i.i.d. random samples, then our conjecture is that empirical distributions using transformed blocks of data are statistically equivalent to the empirical distribution using the transformed random pixels of the same spectral group. We will draw some conclusions from this experiment. Details follow.

Experiment procedures*:* We assembled a *virtual* mosaic of individual HS data consisting of a single spectral group—general terrain (i.e., a mixture of grass, soil, small rocks). Virtual in the sense that although five HS $\left(640 \times 640 \times 120\right)$ data cubes consisting of general terrain, trees areas, and manmade materials were assembled to form an actual mosaic having a total spatial area corresponding to

3200×640, the virtual mosaic did not include radiances from trees and manmade materials—they were masked out from consideration using the segmentation/editing approach described in Subsection 2.4.1. Using a standard randomizer to obtain random locations within the virtual mosaic, we randomly collected 1,698 spectral samples from a single group—this group covered about 60% of the actual mosaic. The random drawing mechanism was analogous to the so-called acceptance-rejection mechanism, where spectral samples that were drawn from off limit areas were rejected. The 1,698 random samples represented about 0.14**%**, or $\dfrac{1698 \cdot 100}{0.6 \cdot (3200 \cdot 640)}$, of the virtual mosaic.

The independently drawn 1,698 spectral samples from this relatively large area of general terrain were transformed using (3.3), (3.5), and (3.7), where $n_2 = 1698$, yielding a sequence as in (3.9). This sequence is considered i.i.d. (IID) in this experiment, since the actual spectral samples (calibrated data) were independently drawn from a relatively wide area consisting of a single spectral group.

Figure 3.1 shows some of the intermediate results required to obtain the empirical distribution of approximately IID single-group spectral samples using our data transformation approach.

Fig. 3.1 (upper left) shows the average spectral sample mean that was estimated using the independently drawn *1,698* spectral samples from a wide area representing a spectral group (terrain), and Fig. 3.1 (upper right) shows this average response after first order differentiation, where the little circles in the plot are the only meaningful points. Fig. 3.1 (lower left) depicts the final data transformed results yielding *1,698* IID, *features* from a single spectral group. And finally Fig. 3.1 (lower

right) shows the *empirical PDF*—this is simply a normalized histogram—using these

IID features as inputs for the estimation step. For convenience, these IID features will

be referred herein as *IID Terrain* and the output features using blocks of data will be

referred to as *Block <number> Terrain*. These labels should emphasize that the data

used for computation are spectral responses from the same spectral group.
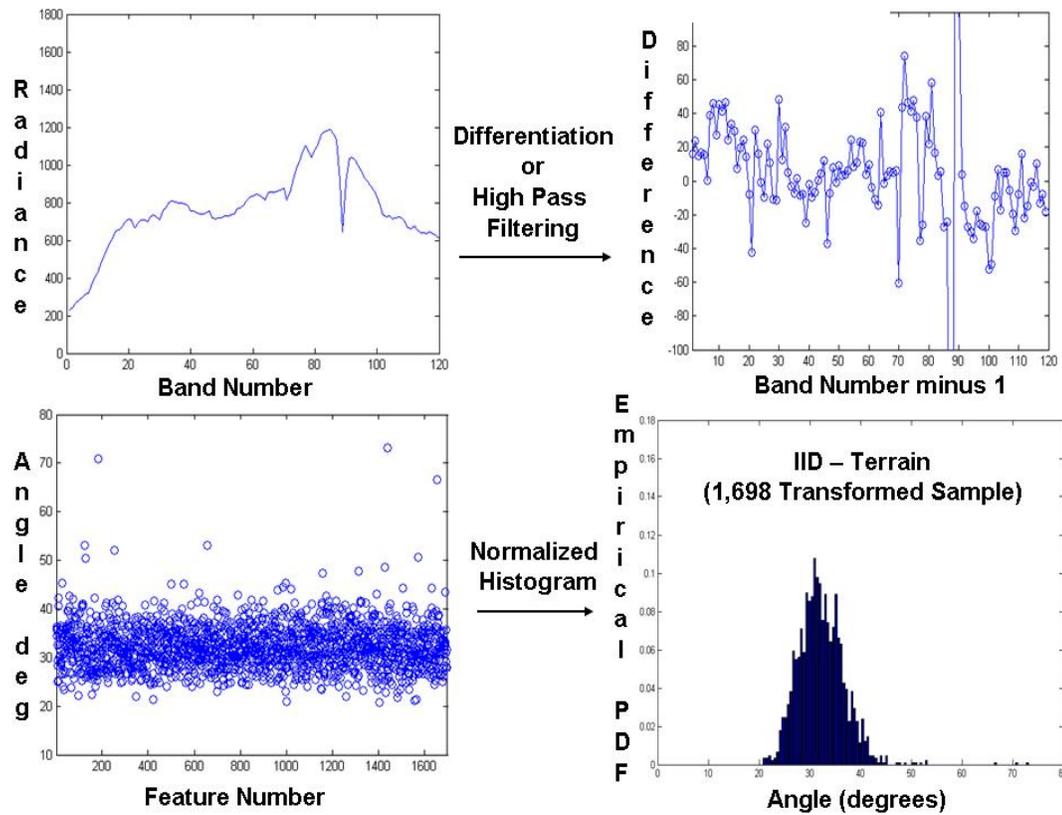
Figure 3.1. The plot at the upper left shows the mean average of the *1,698* independently collected spectral samples of a single group—general terrain. The plot at the upper right illustrates the spectral mean average after differentiation. The plot at the lower left shows the final transformation results (*1,698* univariate features, or sequence) using all of the independently collected spectral samples. The lower right shows the normalized histogram using the *1,698* IID features from a homogeneous terrain, where the vertical axis shows estimated probabilities per angular bin (horizontal axis).

We collected another set of spectral samples from the virtual mosaic, but this

time from a 40×40 block of data, and transformed all spectra $(n_2 = 1600)$ from this

spatial block using (3.3), (3.5), and (3.7)—the same transformation used to generate

IID Terrain. (It is worth noting that the sample sizes of results for IID Terrain and

Block 1 Terrain are significantly larger than the sample size of results obtained for

typical application-based block sizes, about *1,600* versus, for instance, *400*,

respectively; we wanted to study a good approximation to some *ideal* PDF (IID

Terrain) and check for deviations estimating PDFs using terrain spectra from

equivalently large size blocks.) We collected additional blocks of data of the same

size (1600 spectra), and made sure these blocks were sufficiently apart from each

other. After transforming these additional data blocks using the same transformation

approach, we computed histograms independently using each transformed block of

data. These histograms are shown Fig. 3.2 as Block 1 Terrain, Block 2 Terrain,

through Block 8 Terrain.

There are some interesting observations that can be made from the results

shown in Fig. 3.1 and Fig. 3.2. For instance, the histograms for the blocks of

transformed data tend to be reasonably centered with respect to the empirical

distribution of the scene IID transformed spectra (IID Terrain), but their spread

(feature variability) tend to vary depending on whether the blocks of data represented

patches of terrain that are more representative of the overall mix of materials

characterizing the general terrain (grass, soil, rocks), or represented patches of terrain

characterized by a more homogeneous material type (e.g., grass). In the latter case,

the variance of the transformed data seems noticeably smaller than the variance of the
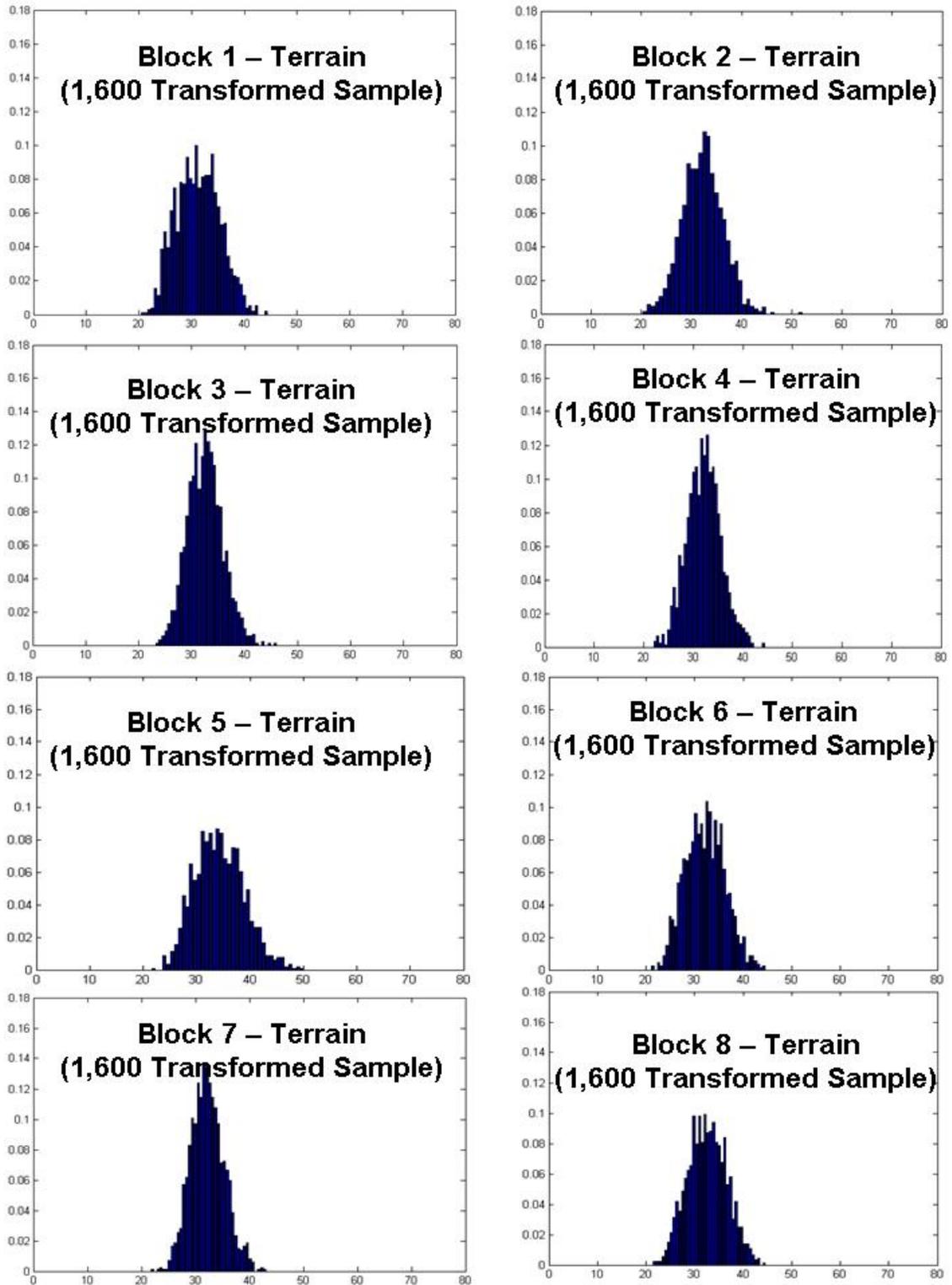
Figure 3.2. Empirical PDFs using 1,600 samples per data block. Data blocks were chosen to be significantly apart from each other.

scene IID transformed data, as one would expect (see, for instance, the empirical PDF of Block 3 and Block 7 in Fig. 3.2 and the empirical PDF of IID Terrain in Fig. 3.1). Quantitative comparisons among these histograms are shown in Fig. 3.3.
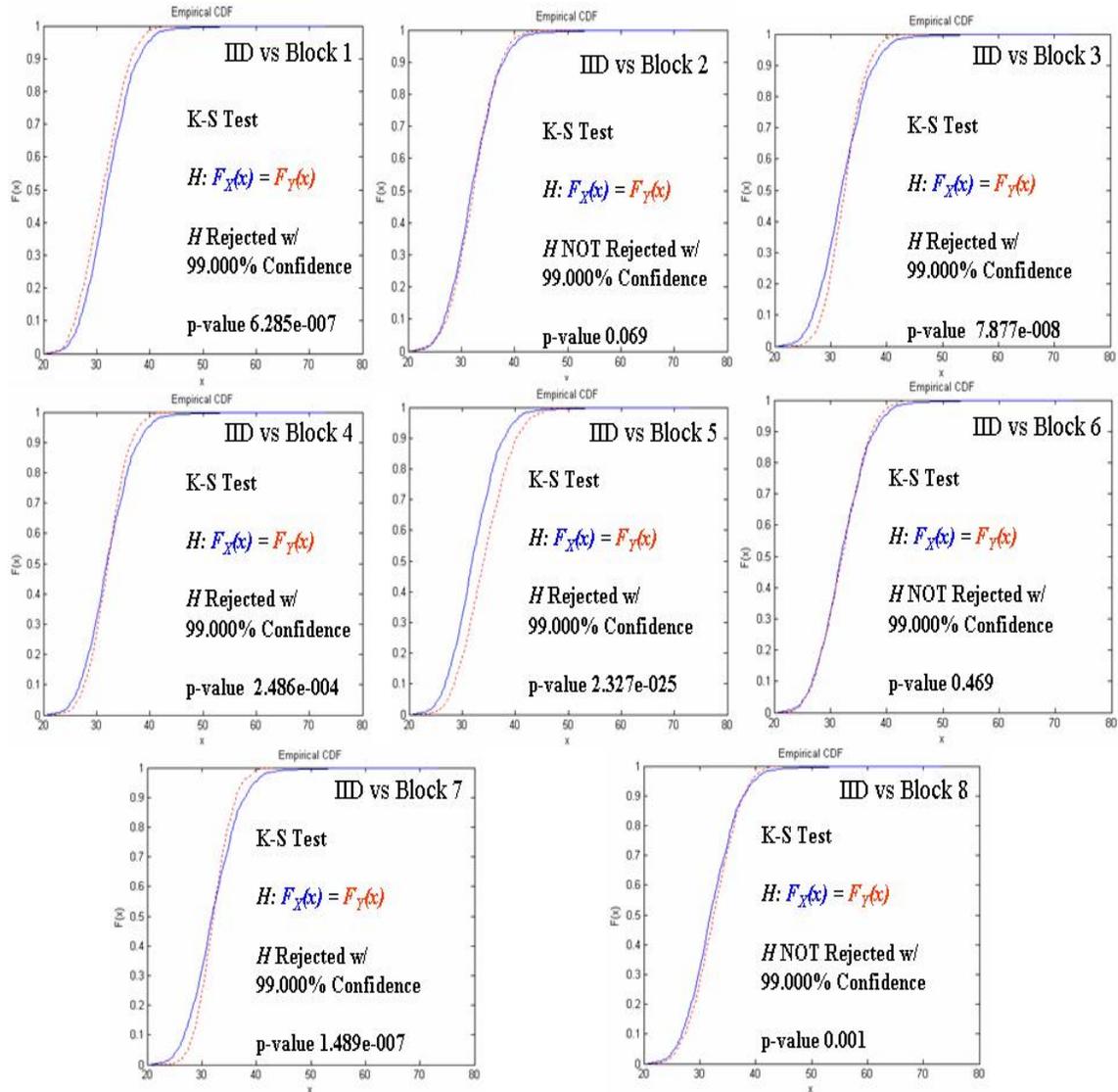
Figure 3.3. Kolmogorov-Smirnov (K-S) test results to check IID assumptions. The cumulative empirical distribution of IID Terrain is represented by solid lines [$F_X(x)$] in the plots, and corresponding cumulative empirical distributions of different Block Terrains are represented by dashed lines [$F_Y(x)$].

We individually compared the empirical PDF of the scene IID transformed

samples to the empirical PDFs of the blocks using the standard K-S test.

For each potential value x, the K-S test compares the proportion of first sample (e.g.,

X) values less than x with proportion of the second sample (e.g., Y) values less than x,

or $\max(F_X(x) - F_Y(x))$, where $F(x)$ denotes the cumulative distribution. The K-S

test results are shown in Fig. 3.3, where the confidence level was set to 99% for each

K-S test.

The difference between normalized histograms of the blocks of data compared

with that of the whole data-cube (IID case) are always significant except in Block 6,

but it is difficult to draw a distinction between the very large and visually obvious

differences in Blocks 1, 3, 5, and 7 as compared with the minor differences in Blocks

2, 4, and 8, because a block of data may not necessarily consist of spectra from a

single homogenous terrain—although it is assumed that it is, and every effort was

made in this experiment to select only blocks of terrain data that appeared to be from

a single terrain. Thus, we settled for quantifying those differences (rather than

explaining them) through the results shown in Figure 3.3.

A *non-rejection* by the K-S test, in this context, means that the corresponding

data transformation approach produces approximately IID sample features, for the

given sample size. Otherwise, it does not produce IID features. The results shown in

Fig 3.3 indicate that some of the data blocks produced approximately IID features,

while other blocks did not. This outcome shows that the confidence one may have in

obtaining IID features from blocks of data using this data transformation approach

cannot be generalized, because it may depend on the location of blocks of data in the

imagery. We could also speculate from these results that blocks of significantly

smaller sample sizes would deviate even more. It is also worth noting that these

results may depend on the homogeneity level of material types in the scene—for

instance, a spectral group (e.g., terrain) often consists of mixtures of different spectral

classes (i.e., terrain may be represented by soil, grass, small rocks, and other

materials).

In summary, this experiment generates a histogram for each 40×40 sampled

block, summarizing the marginal (not joint) distribution of the transformed data

values within each block. Thus this experiment investigates the variability of these

histograms, and compares them with the histogram for the entire large data cube.

The experiment does give information about whether 40×40 contiguous

blocks have marginal distributions (for individual transformed coordinates) that are

sufficiently close to one another. But this experiment tells nothing about the joint

distributions of transformed data values within $n \times n$ windows.

3.4 Summary and Conclusions

We proposed in this chapter a two-step data transformation approach that maps

two calibrated spectral sets from their original space to a subspace, where the

statistical power of detectors can be improved and responses of a given material type

under different illumination conditions are more likely to cluster. The result is then

transformed from this multivariate subspace to a univariate domain, where it is

desired that clusters are preserved. One way to check the plausibility of these desired

features is to conduct simulation experiments that apply different anomaly detection

techniques to transformed data of relevant examples. Chapter 5 discusses such simulation experiments, using data characterization results from Subsection 2.4.

This chapter also presented experiment results that showed that blocks of HS data after data transformation, as described in this chapter, does not satisfy the i.i.d. assumption made for the development of univariate detectors. This outcome is probably also true for other data transformation schemes. The main impact of this outcome is that we might have to develop test cutoffs for our scoring metrics by simulation and sampling (or by some other means) because no theoretical distribution relying on i.i.d. limit theorems can be trusted in the transformed data. Cutoff threshold determination will be addressed in Chapters 5-7.

Chapter 4   Statistical Anomaly Detection

4.1. Introduction

The detectors described in this chapter will be used for a fair comparative study in Chapter 5 and Chapter 6, where it will be shown how the rejection thresholds (cutoffs) are to be obtained, and where it will be clarified under what specific parameter values within the null and alternative hypotheses the desired rejection probabilities are calculated in defining size and power of the test.

Some of the most prominent anomaly detection techniques for HS imagery will be discussed in this chapter. These detectors are based on multivariate techniques (the older detectors) and use two sets of multivariate samples of the form in (3.1) and (3.2) in order to perform their tasks. Multivariate anomaly detectors are the natural choice for HS image data, as discussed in Section 1.3, because a HS sample is multivariate.

This chapter also presents a host of univariate techniques (the new detectors) that, to the best of our knowledge, are applied for the first time to the HS anomaly detection problem. These detectors take as input two sets of univariate samples of the form in (3.9) and (3.10) in order to perform their tasks.

Note that the older anomaly detectors enter the comparison study only as applied to the calibrated K-dim HS data, while the new detectors including all those in Sec. 4.3 operate only on the transformed data.

The remainder of this chapter is organized as follows: Section 4.2 presents a brief discussion on multivariate normal based statistical hypothesis testing

59

(Subsection 4.2.1), describes the industry standard for HS anomaly detection

(Subsection 4.2.1), and presents alternative multivariate anomaly detection techniques

(Subsection 4.2.2).

Section 4.3 introduces univariate techniques for HS anomaly detection, including a

semiparametric model (Subsection 4.3.1) and other alternative data combining

techniques (Subsection 4.3.2). Section 4.4 summarizes this chapter.

4.2. Multivariate Techniques

Since the most popular older detector is based on a multivariate statistical

technique, we will start with a brief description on statistical hypothesis testing and

continues with a discussion on target/anomaly detection based on multivariate normal

distributions, followed by some of the additional alternative multivariate techniques.

In particular, the following multivariate detectors are discussed: Reed-Xi (RX)

detector [14], the kernel-based RX (KRX) detector [15], Fisher's linear discriminant

(FLD) detector [16], dominant principal component (DPC) detector [16] and Eigen

separation transform (EST) detector [16]. These techniques, or variants of them,

arguably represent a list of the most distinct approaches for HS anomaly detection.

But conspicuously missing from this list are techniques based on Markov Chain

Monte Carlo (MCMC). We found only one specific MCMC based anomaly detector

in the open literature [24], but we excluded it from this effort because its performance

was shown [24] to be comparable, not improved, to that of the RX detector.

4.2.1. Statistical Hypothesis Testing

A common problem in science is the empirical verification or rejection of a hypothesis concerning a population. In essence, a statistical hypothesis test formalizes the various actions that can be taken and introduces explicitly a numerical measure of the consequence of each action for a given state of Nature. One may interpret this formality as a process, where one out of two competing hypotheses regarding the population of a random variable will be chosen as the most likely hypothesis. In general, one defines two hypotheses (a simple or composite null hypothesis and a simple or composite alternative hypothesis [25]) and derives a test statistic under the conditions of the simple or composite null hypothesis to show that under this hypothesis the test statistic has a known behavior, i.e., it is controlled by a known probability distribution function. The test statistic then is itself a random variable defined as a function of the random variables that comprise a random sample.

Since all the detectors used for this work require a two-sample data structure, a simple hypothesis, in this context, means that the null or alternative probability law for the two-sample data is completely specified including all parameter values. In the same context, a composite hypothesis means that the null or alternative probability law corresponds to a family of possible distributions with parameter values not fully known in advance.

We discuss next a target detection test under the multivariate normal assumption using the HS imagery.

4.2.1.1. Multivariate Normal Target Detection

In principle, target detection problems can be formalized making use of a null hypothesis and an alternative hypothesis, simple or composite. These hypothesis tests attempt to answer whether the observed samples correspond to target data or non-target data. Depending on how much is known a priori about the target(s), implementation of this test may take different forms. For instance, if one uses the likelihood ratio method [25] to design a target detector and has a priori knowledge that the null and alternative hypothesis distributions are multivariate normal, then, for a single background group (e.g., G1—see Subsection 2.4.1) and a particular target (G4—see Subsection 2.4.1), the null and alternative hypotheses are:

$$H_0 : \mathbf{x}_{rc} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \quad \text{Target Absent}$$
$$H_1 : \mathbf{x}_{rc} \sim N(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4) \quad \text{Target Present} \tag{4.1}$$

where $\mathbf{x}_{rc} \in \mathbf{R}^K$ is a random column vector at the $(r,c)^{th}$ pixel location in $\mathbf{X}$ [see (2.4)], the K dimensional mean column vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_4$ are different from each other, and the $K \times K$ dimensional covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_4$ are also different from each other.

If parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_4)$ are completely known a priori, then (4.1) are simple hypotheses and the natural logarithm of the likelihood ratio detector [4] is known as the quadratic discriminant function,

$$d_Q^{(rc)} = (\mathbf{x}_{rc} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_1^{-1} (\mathbf{x}_{rc} - \boldsymbol{\mu}_1) - (\mathbf{x}_{rc} - \boldsymbol{\mu}_4)^t \boldsymbol{\Sigma}_4^{-1} (\mathbf{x}_{rc} - \boldsymbol{\mu}_4) , \tag{4.2}$$

where $\left\{d_Q^{(rc)}\right\}_{r=1,c=1}^{R,C}$ are non-negative scalars and $\mathbf{\Sigma}^{-1}$ is the inverse of matrix $\mathbf{\Sigma}$.

When the covariance of the two hypothesis distributions are equal $\left(\mathbf{\Sigma}_1 = \mathbf{\Sigma}_4 = \mathbf{\Sigma}\right)$, the quadratic detector simplifies to the matched filter [4],

$$d_{MF}^{(rc)} = \eta \mathbf{x}_{rc}^t \mathbf{\Sigma}^{-1}\left(\mathbf{\mu}_4 - \mathbf{\mu}_1\right), \tag{4.3}$$

where η is a normalizing constant.

In actual applications, however, the exact distributions are not known a priori. Under these conditions, the mean and covariance of each hypothesis normal distribution can be replaced with their maximum likelihood (ML) estimates. This is known as the generalized likelihood ratio (GLR) detector [4]. When ML estimates are used in (4.3), enough data points are required to estimate the mean and covariance for both hypothesis distributions. For the sake of this discussion, let assume that enough data points are available a priori, then $\hat{\mathbf{\mu}}_1$, $\hat{\mathbf{\mu}}_4$, and $\hat{\mathbf{\Sigma}}$ could be computed and stored in a library. Using this library, unknown vectors $\left\{\mathbf{x}_{rc}\right\}_{r=1,c=1}^{R,C}$ in $\mathbf{X}$ could be tested according to (4.1) via the following adaptive matched filter:

$$d_{AMF}^{(rc)} = \eta \mathbf{x}_{rc}^t \hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\mu}}_4 - \hat{\mathbf{\mu}}_1\right), \tag{4.4}$$

where, $\hat{\mathbf{\mu}}_1$ is the estimator of $\mathbf{\mu}_1$, $\hat{\mathbf{\mu}}_4$ is the estimator of $\mathbf{\mu}_4$, and $\hat{\mathbf{\Sigma}}$ is the estimator of $\mathbf{\Sigma}$.

However, using (4.4) to test $\{\mathbf{x}_{rc}\}_{r=1, c=1}^{R, C}$ in $\mathbf{X}$ might not be reliable, because the ML estimates $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_4, \hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_4)$, which are obtained from a pre-stored spectral library, are often found not to be robust for all kinds of environmental and/or illumination conditions. This problem leads to the need for anomaly detection, where all data points in $\mathbf{X}$ would be tested using reference data directly from $\mathbf{X}$ (not from a spectral library), such that, any anomalous object in the scene would be declared a candidate target.

4.2.1.2. Multivariate Normal Anomaly Detection

An anomaly detector that is based on the family of normal distributions is derived from the likelihood ratio test [25], where the exact shape $\boldsymbol{\Sigma}$ and location $\boldsymbol{\mu}$ of the target hypothesis H_1 distribution are not known. For simplicity, let's assume again that background spectra belong to a single group (G1) having a K-dimensional mean column vector $\boldsymbol{\mu}_1$. Under the assumption of normally distributed hypotheses with equal covariances $(\boldsymbol{\Sigma})$, the binary hypothesis test is

$$
\begin{aligned}
H_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_1 \\
H_1 &: \boldsymbol{\mu} \neq \boldsymbol{\mu}_1
\end{aligned}
\tag{4.5}
$$

where the column vector $\boldsymbol{\mu} \in \mathbf{R}^K$ is the mean of a test pixel $\mathbf{x}_{rc} \in \mathbf{R}^K$ in $\mathbf{X}$, see (2.4).

If $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}$ are completely known, the likelihood ratio test simplifies to the Mahalanobis distance metric [4],

$$d_{MD}^{(rc)} = \left(\mathbf{x}_{rc} - \boldsymbol{\mu}_1\right)^t \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_{rc} - \boldsymbol{\mu}_1\right), \tag{4.6}$$

and the adaptive anomaly detector using the ML estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ is

$$d_{AMD}^{(rc)} = \left(\mathbf{x}_{rc} - \hat{\boldsymbol{\mu}}\right)^t \hat{\boldsymbol{\Sigma}}^{-1} \left(\mathbf{x}_{rc} - \hat{\boldsymbol{\mu}}\right), \tag{4.7}$$

where the computations of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ depend on the sensor's viewing perspective recording $\mathbf{X}$ [for instance, for top view imagery, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are estimates using pixels (spectra) surrounding $\mathbf{x}_{rc}$ (see Chapter 6); for ground view imagery, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are estimates using spectra from a spectral library]. Pixel $\mathbf{x}_{rc}$ would be declared an anomaly when scalar $d_{AMD}^{(rc)}$ is greater than a high cutoff threshold.

The adaptive anomaly detector in (4.7) forms the basis for the so called Reed-Xiaoli (RX) algorithm, which is discussed next.

4.2.1.3.   Reed-Xi (RX) Anomaly Detector

Based on the material discussed in Subsection 4.2.1.1 and Subsection 4.2.1.2, a fully adaptive multiband spectral detector was proposed by Reed and Yu in [14]. This detector is a generalized version of the adaptive spectral matched filter, where the problem was formulated to detect objects of a known spatial pattern, but unknown spectral distribution, against a clutter background that is assumed to be normally distributed with unknown parameters. If all the assumptions are satisfied (i.e., spectra of clutter background are i.i.d. and normal), then this detection test has a

constant false alarm rate (CFAR) property over the detector's response for clutter background.

This detector was employed by the DARPA MUSIC program [26] to detect military vehicles in an intense clutter background, and this visibility made the RX algorithm the industry standard for anomaly detection in the target community. Since then, The RX anomaly detector has become a baseline approach for comparison purposes against competing anomaly detection approaches.

Favorable claims have been made, see [26] and [27], suggesting that the RX anomaly detector is robust when used to detect spectral differences between a block of data in $\mathbf{X}$ —see the $n \times n$ window in (2.8) —and a reference spectra set. But for an anomaly detector to be truly robust, or even effective in all types of realistic scenarios (see Chapter 7, where real HS data are used), it must be able to handle samples from all kinds of data, including from the ones shown in the null and alternative hypotheses (2.11) and (2.12), respectively. (Recall that (2.11) and (2.12) will be used as baseline hypotheses for the simulation experiments discussed in Chapter 5.)

Using the two sample implementation method described in Subsection 2.2.3, where $\mathbf{W}_2 = \left(\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2} \right)$, column vectors $\left\{ \mathbf{y}_{2u} \right\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right)$, and column vectors $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n_1} \in \mathbf{R}^K$, a popular version of the RX anomaly detector is

$$Z_{RX}^{(ij)} = \frac{n_1 n_2}{n_1 + n_2} \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 \right)^{\mathrm{t}} \left(\hat{\mathbf{\Sigma}}_2^{-1} \right) \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 \right), \qquad (4.8)$$

where $Z_{RX}^{(ij)} \geq 0.0$, $\overline{\mathbf{y}}_1 = n_1^{-1}\sum_{u=1}^{n_1}\mathbf{y}_{1u}$, $\overline{\mathbf{y}}_2 = n_2^{-1}\sum_{u=1}^{n_2}\mathbf{y}_{2u}$,

$$\hat{\mathbf{\Sigma}}_2 = (n_2 - 1)^{-1}\sum_{u=1}^{n_2}(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)^t , \ (i = 1,\cdots, R - n - 1) \text{ and } (j = 1,\cdots, C - n - 1)$$

index the left upper corner pixel of an $n \times n$ window in $\mathbf{X}$ [see, for instance, (2.8)

and (2.9)].

Note that in order to test the entire $\mathbf{X}$, all $\{Z_{RX}^{(ij)}\}_{i=1,\ j=1}^{R-n-1,\ C-n-1} \in \mathbf{R}$ must be

computed using (4.8), where $\mathbf{W}_1$ and $\mathbf{W}_2$ are obtained as described in Subsection

2.2.3. A *2-dim* output surface can then be constructed using $\{Z_{RX}^{(ij)}\}_{i=1,\ j=1}^{R-n-1,\ C-n-1}$, or

$$\mathbf{Z}_{RX} = \begin{bmatrix} Z_{RX}^{(11)}, & Z_{RX}^{(12)}, & \cdots, & Z_{RX}^{[1(C-n-1)]} \\ Z_{RX}^{(21)}, & Z_{RX}^{(22)}, & \cdots, & Z_{RX}^{[2(C-n-1)]} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{RX}^{[(R-n-1)1]}, & Z_{RX}^{[(R-n-1)2]}, & \cdots, & Z_{RX}^{[(R-n-1)(C-n-1)]} \end{bmatrix} \tag{4.9}$$

where $\mathbf{Z}_{RX} \in \mathbf{R}^{(R-n-1)\times(C-n-1)}$ has a smaller spatial area than $\mathbf{X}$'s $R \times C$ spatial area.

Output surfaces having the form of (4.9) will be shown later in Chapter 6 (testing

results for top view imagery) and Chapter 7 (testing results for ground view imagery)

for different ways of obtaining $\mathbf{W}_1$ and $\mathbf{W}_2$.

4.2.1.4. Kernel RX (KRX) Algorithm

The RX detector does not take into account the higher order relationships

between the spectral bands at different wavelengths. The nonlinear relationships

between different spectral bands within the target or clutter spectral signature were

exploited in [15] using a kernel-based version of the RX model. The authors named

this approach: the kernel RX (KRX) algorithm.

An interpretation of the KRX algorithm is that it extends the utility of the RX

algorithm from a lower dimensional data space to a higher dimensional nonlinear

feature space by applying a well known kernel trick (see, for instance, [15]) in order

to *kernelize* the corresponding generalized likelihood ratio test expression of the RX

approach. The GLR test expression of the kernel RX is similar to the RX approach,

but every term in the expression is in kernel form, which can be readily calculated in

terms of the input data in its original data space.

The notion of applying nonlinear kernels as a means to extract features from

data is not new. The most prominent algorithm using this application is the well

known support vector machine proposed by Vapnik [28]. Many other kernel-based

versions of existing algorithms have been proposed in the literature, including kernel

PCA [17] and kernel FLD [29]. The authors of the KRX detector, however, were the

first ones to present to the HS research community the kernelized version of the RX

algorithm.

The implicit model of the KRX method is very different from the RX method,

namely that the data values are multivariate normal not as HS measurements, but only

when those measured values are nonlinearly embedded in some higher dimensional

space. Using the two sample implementation method described in Subsection 2.2.3,

where $\mathbf{W}_2 = (\mathbf{y}_{21}, \cdots, \mathbf{y}_{2n_2})$, $\{\mathbf{y}_{2u}\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = (\mathbf{y}_{11}, \cdots, \mathbf{y}_{1n_1})$, and $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$,

the KRX anomaly detector can be compactly represented by

$$Z_{KRX}^{(ij)} = (\mathbf{K}_{\bar{\mathbf{y}}_1} - \mathbf{K}_{\bar{\mathbf{y}}_2})^t \mathbf{K}_2^{-1} (\mathbf{K}_{\bar{\mathbf{y}}_1} - \mathbf{K}_{\bar{\mathbf{y}}_2}), \qquad (4.10)$$

where $Z_{KRX}^{(ij)} \geq 0.0$, $\mathbf{K}_{\bar{\mathbf{y}}_1} = \mathbf{K}(\bar{\mathbf{y}}_1, \mathbf{W}_2)$ is a kernel-function based vector that uses as

input $\bar{\mathbf{y}}_1 = n_1^{-1} \sum_{u=1}^{n_1} \mathbf{y}_{1u}$ and $\mathbf{W}_2$, $\mathbf{K}(\mathbf{v}_1, \mathbf{v}_2)$ represents the dot product between $\mathbf{v}_1$ and

$\mathbf{v}_2$ nonlinearly mapped to a higher dimensional space (the mapping function will be

shown later), $\mathbf{K}_{\bar{\mathbf{y}}_2} = \mathbf{K}(\bar{\mathbf{y}}_2, \mathbf{W}_2)$ is a kernel-function based vector applying the same

kernel function but using instead the dot product between $\bar{\mathbf{y}}_2 = n_2^{-1} \sum_{u=1}^{n_2} \mathbf{y}_{2u}$ and $\mathbf{W}_2$,

$\mathbf{K}_2^{-1}$ is the inverse of $\mathbf{K}_2 = \mathbf{K}(\mathbf{W}_2, \mathbf{W}_2)$ (which represents the dot product between

$\mathbf{W}_2$ and itself using the same kernel function), $\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}$ and $\mathbf{W}_2 \in \mathbf{R}^{K \times n_2}$.

The rationale for using $\mathbf{K}_2^{-1}$ as the normalizing matrix in (4.10) is based on the

properties of the kernel PCA, which is explained, for instance, in [15].

Finally, the *atomic* kernel function used to implement the KRX detector in

this effort was the well known Gaussian (radial basis function) RBF kernel [15].

To test the entire $\mathbf{X}$, all

$$\left\{ Z_{KRX}^{(ij)} \right\}_{i=1, \, j=1}^{R-n-1, \, C-n-1} \in \mathbf{R} \qquad (4.11)$$

must be computed, where $\mathbf{W}_1$ and $\mathbf{W}_2$ are obtained as described in Subsection 2.2.3.

A *2-dim* output surface can be constructed using $\left\{ Z_{KRX}^{(ij)} \right\}_{i=1, \, j=1}^{R-n-1, \, C-n-1}$, or

$$\mathbf{Z}_{KRX} = \begin{bmatrix} Z_{KRX}^{(11)}, & \cdots, & Z_{KRX}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{KRX}^{[(R-n-1)1]}, & \cdots, & Z_{KRX}^{[(R-n-1)(C-n-1)]} \end{bmatrix} \tag{4.12}$$

4.2.2. Alternative Multivariate Anomaly Detection

This subsection describes three additional multivariate based techniques for HS anomaly detection: Fisher's linear discriminant and two eigen based approaches.

4.2.2.1. Fisher's Linear Discriminant (FLD)

Fisher's linear discriminant analysis has become a standard technique for detection problems involving samples from different object classes. It projects the original high dimensional data onto a low dimensional space, where all the classes are well separated by maximizing the *Rayleigh quotient*, i.e., the ratio of between-class scatter matrix determinant to within-class scatter matrix determinant. The application of the FLD detector to HS imagery has been investigated for anomaly detection [16] and for object classification [27], where a classification algorithm was derived based on FLD to force different classes to be along different directions in a low dimensional space. Multi-object classification is beyond the scope of this dissertation; hence, our focus will be limited to adapting FLD to a two class problem in HS imagery.

Using $\mathbf{W}_2 = \left(\mathbf{y}_{21}, \cdots, \mathbf{y}_{2n_2}\right)$, $\left\{\mathbf{y}_{2u}\right\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \mathbf{y}_{1n_1}\right)$, and

$\left\{\mathbf{y}_{1h}\right\}_{h=1}^{n_1} \in \mathbf{R}^K$ as described in Subsection 2.2.3, a version of FLD for the two-class

(anomaly or not anomaly) problem is shown below:

$$Z_{FLD}^{(ij)} = \left| \left(\mathbf{E}_{\mathbf{S}_B \mathbf{S}_W^{-1}}^t\right)\left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2\right) \right|, \tag{4.13}$$

where $\overline{\mathbf{y}}_1 = n_1^{-1}\sum_{u=1}^{n_1}\mathbf{y}_{1u}$, $\overline{\mathbf{y}}_2 = n_2^{-1}\sum_{u=1}^{n_2}\mathbf{y}_{2u}$, $\left| \bullet \right|$ denotes the absolute value operator, and

$\mathbf{E}_{\mathbf{S}_B \mathbf{S}_W^{-1}}^t \in \mathbf{R}^K$ is the transposed highest energy eigenvector (column vector) from the

principal component decomposition of $\mathbf{S}_B \mathbf{S}_W^{-1}$, such that

$$\mathbf{S}_W = \sum_{u=1}^{n_1}(\mathbf{y}_{1u} - \overline{\mathbf{y}}_1)(\mathbf{y}_{1u} - \overline{\mathbf{y}}_1)^t +$$
$$\sum_{u=1}^{n_2}(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)^t, \tag{4.14}$$

$$\mathbf{S}_B = \sum_{u=1}^{n_1}(\mathbf{y}_{1u} - \overline{\overline{\mathbf{y}}})(\mathbf{y}_{1u} - \overline{\overline{\mathbf{y}}})^t +$$
$$\sum_{u=1}^{n_2}(\mathbf{y}_{2u} - \overline{\overline{\mathbf{y}}})(\mathbf{y}_{2u} - \overline{\overline{\mathbf{y}}})^t, \tag{4.15}$$

where $\overline{\overline{\mathbf{y}}} = \left(n_1 + n_2\right)^{-1}\left(\sum_{u=1}^{n_1}\mathbf{y}_{1u} + \sum_{h=1}^{n_2}\mathbf{y}_{2h}\right)$.

After testing the entire $\mathbf{X}$, all $\left\{ Z_{FLD}^{(ij)} \right\}_{i=1,\ j=1}^{R-n-1,\ C-n-1} \in \mathbf{R}$ are computed, where $\mathbf{W}_1$ and $\mathbf{W}_2$ are obtained as described in Subsection 2.2.3. The corresponding output surface $\mathbf{Z}_{FLD}$ can be constructed as follows

$$\mathbf{Z}_{FLD} = \begin{bmatrix} Z_{FLD}^{(11)}, & \cdots, & Z_{FLD}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{FLD}^{[(R-n-1)1]}, & \cdots, & Z_{FLD}^{[(R-n-1)(C-n-1)]} \end{bmatrix}. \tag{4.16}$$

4.2.2.2.   Dominant Principal Component (DPC)

The DPC technique is based on a basic general principle, that data are projected from their original high dimensional space onto a significantly lower dimensional space (in our case, only one dimension) using a criterion that promotes highest sample variability within each domain in this lower dimensional space. This technique has been also adapted to HS anomaly detection and yielded some good performances on real HS data (see, for instance, [16]).

Using $\mathbf{W}_2 = \left(\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2} \right)$, $\left\{ \mathbf{y}_{2u} \right\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right)$, and $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n_1} \in \mathbf{R}^K$ as discussed in Subsection 2.2.3, this technique can be adapted to the two-class (anomaly or not anomaly) problem as follows:

$$Z_{DPC}^{(ij)} = \left| \mathbf{E}_2^t \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 \right) \right|, \tag{4.17}$$

where $\overline{\mathbf{y}}_1 = n_1^{-1} \sum\limits_{u=1}^{n_1} \mathbf{y}_{1u}$, $\overline{\mathbf{y}}_2 = n_2^{-1} \sum\limits_{u=1}^{n_2} \mathbf{y}_{2u}$, and $\mathbf{E}_{\hat{\Sigma}_2}^t \in \mathbf{R}^K$ is the transposed highest

energy eigenvector (column vector) from the principal component decomposition of

$$\hat{\mathbf{\Sigma}}_2 = (n_2 - 1)^{-1} \sum\limits_{u=1}^{n_2} (\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)^t \; .$$

An output surface $\mathbf{Z}_{DPC}$ using (4.17) can be constructed as follows

$$\mathbf{Z}_{DPC} = \begin{bmatrix} Z_{DPC}^{(11)}, & \cdots, & Z_{DPC}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{DPC}^{[(R-n-1)1]}, & \cdots, & Z_{DPC}^{[(R-n-1)(C-n-1)]} \end{bmatrix} . \tag{4.18}$$

4.2.2.3.   Eigen-Separation Transform (EST)

The EST detector is a variation of the DPC detector, see [16]. Using

$\mathbf{W}_2 = (\mathbf{y}_{21}, \cdots, \; \mathbf{y}_{2n_2})$, $\{\mathbf{y}_{2u}\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = (\mathbf{y}_{11}, \cdots, \; \mathbf{y}_{1n_1})$, and $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$, as

described in Subsection 2.2.3, the EST anomaly detector can be expressed as

$$Z_{EST}^{(ij)} = \left| \mathbf{E}_{\Delta\Sigma}^t \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 \right) \right| , \tag{4.19}$$

where $\overline{\mathbf{y}}_1 = n_1^{-1} \sum\limits_{u=1}^{n_1} \mathbf{y}_{1u}$, $\overline{\mathbf{y}}_2 = n_2^{-1} \sum\limits_{u=1}^{n_2} \mathbf{y}_{2u}$, and $\mathbf{E}_{\Delta\Sigma}^t \in \mathbf{R}^K$ is the transposed highest

energy eigenvector from the principal component decomposition of $\mathbf{\Delta\Sigma} = \hat{\mathbf{\Sigma}}_1 - \hat{\mathbf{\Sigma}}_2$,

such that, $\hat{\mathbf{\Sigma}}_1 = \left(n_1 - 1\right)^{-1} \sum_{u=1}^{n_1} (\mathbf{y}_{1u} - \overline{\mathbf{y}}_1)(\mathbf{y}_{1u} - \overline{\mathbf{y}}_1)^t$ and

$$\hat{\mathbf{\Sigma}}_2 = \left(n_2 - 1\right)^{-1} \sum_{u=1}^{n_2} (\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)(\mathbf{y}_{2u} - \overline{\mathbf{y}}_2)^t .$$

An output surface $\mathbf{Z}_{EST}$ using (4.19) can be constructed as follows

$$\mathbf{Z}_{EST} = \begin{bmatrix} Z_{EST}^{(11)}, & \cdots, & Z_{EST}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{EST}^{[(R-n-1)1]}, & \cdots, & Z_{EST}^{[(R-n-1)(C-n-1)]} \end{bmatrix}. \tag{4.20}$$

4.3. Univariate Techniques

All the univariate techniques presented in this section use results from the data transformation presented in Chapter 3. These techniques range from parametric, semiparametric, to nonparametric. These univariate techniques will be sometimes referred to in this dissertation as a two-step approach for HS anomaly detection, because two multivariate data sets are first mapped to two univariate series, then a resultant scalar is produced from these univariate series via a score metric.

The remainder of this section is organized as follows: Subsection 4.2.1 discusses a semiparametric model and Subsection 4.2.2 discusses alternative parametric and nonparametric techniques, including the well known analysis of variance.

4.3.1. Semiparametric (SemiP) Anomaly Detection

A semiparametric model and its application to anomaly detection is discussed in this subsection. Using the same two sample implementation method described in

Subsection 2.2.3, where $\mathbf{W}_2 = (\mathbf{y}_{21}, \cdots, \mathbf{y}_{2n_2})$, $\{\mathbf{y}_{2u}\}_{u=1}^{n_2} \in \mathbf{R}^K$, $\mathbf{W}_1 = (\mathbf{y}_{11}, \cdots, \mathbf{y}_{1n_1})$,

and $\{\mathbf{y}_{1h}\}_{h=1}^{n_1} \in \mathbf{R}^K$, these samples will be transformed, as described in Section 3.2, to

$x_2 = (x_{21}, \cdots, x_{2n_2}) \sim f_2(x)$ and $x_1 = (x_{11}, \cdots, x_{1n_2}) \sim f_1(x)$, where $0^0 \leq \{x_{1u}\}_{u=1}^{n_2} \leq 90^0$,

$0^0 \leq \{x_{2u}\}_{u=1}^{n_2} \leq 90^0$, and $f_2(x)$ and $f_1(x)$ are unknown joint PDFs.

In order to simplify the anomaly detection problem using mapped data, let x_{1u}

and x_{2u} $(u = 1, \cdots, n_2)$ be i.i.d. random variables controlled by unknown marginal

PDFs g_0 and g_1, respectively, or

$$
\begin{aligned}
x_2 &= (x_{21}, \ldots, x_{2n_2}) \text{ i.i.d. } \sim g_0(x) \\
x_1 &= (x_{11}, \ldots, x_{1n_2}) \text{ i.i.d. } \sim g_1(x)
\end{aligned}
\tag{4.21}
$$

where $\{x_{2u}\}_{u=1}^{n_2}$ and $\{x_{1u}\}_{u=1}^{n_2}$ are assumed to be independent from each other.

If we assume further that these marginal distributions are exponentially related

as

$$
\frac{g_1(x)}{g_0(x)} = \exp(\alpha + \beta x),
\tag{4.22}
$$

but are otherwise unrestricted, then since g_1 is a density, $\beta = 0$ must imply that $\alpha = 0$. Thus, α merely functions as a normalizing parameter. Notice also in (4.22) that if $\beta = 0$ then $\{x_{2u}\}_{u=1}^{n_2}$ and $\{x_{1u}\}_{u=1}^{n_2}$ must belong to the same population (i.e., $g_1 = g_0$).

Using this fact, an anomaly detection test statistic can be designed by checking the

following null (H_0) and alternative (H_1) hypotheses:

75

$$H_0: \quad \beta = 0 \quad \left(\frac{g_1}{g_0} = 1.0\right) \qquad \text{anomaly absent}$$

$$H_1: \quad \beta \neq 0 \quad \left(\frac{g_1}{g_0} = \exp(\alpha + \beta x)\right) \quad \text{anomaly present} \qquad \qquad (4.23)$$

The following test statistic can be obtained [30][31] from (4.21), (4.22), and (4.23):

$$Z^{(ij)}_{SemiP} = n\rho(1+\rho)^{-2}\hat{\beta}^2\hat{v}^2, \qquad\qquad (4.24)$$

where $i = 1, \cdots, R - n - 1$ and $j = 1, \cdots, C - n - 1$ index the left upper corner pixel of an

$n \times n$ window in $\mathbf{X}$ [see (2.8) and (2.9)], $t = (x_{21}, ..., x_{2n_2}, x_{11}, ..., x_{1n_2}) \equiv (t_1, ..., t_{n_0})$;

$\hat{v}^2 = \sum_i t_i^2 \hat{\tilde{g}}_0(t_i) - \left(\sum_i t_i \hat{\tilde{g}}_0(t_i)\right)^2$; $n_0 = n_2 + n_2$; using a profiling procedure (see

[30][31]), $\hat{\tilde{g}}_0(t_i) = \dfrac{1}{n_2}\dfrac{1}{1 + \rho\exp(\hat{\alpha} + \hat{\beta}t_i)}$ estimates $\tilde{g}_0(t_i) = \dfrac{1}{n_0}\dfrac{1}{1 + \rho\exp(\alpha + \beta t_i)}$,

which in turn estimates g_0 ; $\left(\hat{\alpha}, \hat{\beta}\right)$ are the ML (maximum likelihood) estimators of

(α, β) ; and $\rho(1+\rho)^{-2}$ is a constant related to sample sizes of x_2 and x_1, see

[30][31].

The test statistic in (4.24) will be referred herein as the SemiP test statistic,

and the two step approach—data mapping (see Section 2.2.3) and application of the

SemiP test statistic—will be referred herein as the SemiP anomaly detector. (For

convenience, we may also refer to results produced by the SemiP anomaly detector as $Z_{SemiP}^{(ij)}$.)

Using (4.24), the entire $\mathbf{X}$ will be used for testing, where $\mathbf{W}_1$ and $\mathbf{W}_2$ may be obtained as described in Subsection 2.2.3. A *2-dim* output surface can be constructed from $\{Z_{SemiP}^{(ij)}\}_{i=1,\,j=1}^{R-n-1,\,C-n-1}$,

$$\mathbf{Z}_{SemiP} = \begin{bmatrix} Z_{SemiP}^{(11)}, & \cdots, & Z_{SemiP}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{SemiP}^{[(R-n-1)1]}, & \cdots, & Z_{SemiP}^{[(R-n-1)(C-n-1)]} \end{bmatrix}. \qquad (4.25)$$

Examples of $\mathbf{Z}_{SemiP} \in \mathbf{R}^{(R-n-1)\times(C-n-1)}$ are shown in Chapter 6 (testing simulated top view data cubes) and Chapter 7 (testing real ground view HS data cubes) for different ways of obtaining $\mathbf{W}_2$. (Recall that $\mathbf{W}_1$ is always obtained using a small window at the $(ij)^{th}$ location in the imagery.)

4.3.2. Alternative Univariate Methods for Anomaly Detection

Motivated by the advantage and disadvantage of using semiparametric inference for anomaly detection, we propose to use three alternative univariate scoring metrics. These metrics also combine samples in order to perform their functions, but, differently from the SemiP anomaly detector, they do not require parameter initialization.

The scoring metrics described in this subsection are applied for the first time to the problem of HS anomaly detection.

4.3.2.1. Functional Approximation of SemiP (AsemiP)

We present a nonparametric score metric that is free from parameter initialization and can functionally behave like the semiparametric test statistic in (4.24) when introduced with two sequences of univariate data.

Starting with two multivariate samples $\mathbf{W}_2 = (\mathbf{y}_{21}, \cdots, \mathbf{y}_{2n_2})$ and $\mathbf{W}_1 = (\mathbf{y}_{11}, \cdots, \mathbf{y}_{1n_1})$, these samples are transformed, as described in Section 3.2, to univariate sequences $x_2 = (x_{21}, \cdots, x_{2n_2})$ and $x_1 = (x_{11}, \cdots, x_{1n_2})$. We propose to combine these sequences in the following way in order to decide whether x_1 is anomalous to x_2:

$$Z_{AsemiP}^{(ij)} = \left(n_2^{-1} + n_2^{-1}\right)^{-1} \frac{\hat{\tilde{\beta}}^2}{S^2}\left(\frac{S_{union}^2}{S_2^2}\right),$$

(4.30)

where,

$$t = \left(x_{21}, \cdots, x_{2n_2}, x_{11}, \cdots, x_{1n_2}\right) = \left(t_1, \ldots, t_{n_0}\right), \ n_0 = n_2 + n_2,$$

(4.31)

$$S_{union}^2 = (n-1)^{-1} \sum_{i=1}^{n} (t_i - \bar{t})^2,$$

(4.32)

$$\bar{t} = n^{-1} \sum_{i=1}^{n} t_i,$$

(4.33)

$$S^2 = \frac{(n_2 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_2 - 1) + (n_2 - 1)}, \qquad (4.34)$$

$$S_1^2 = \frac{\sum_{i=1}^{n_2}(x_{1i} - \bar{x}_1)^2}{n_2 - 1}, \qquad (4.35)$$

$$\bar{x}_1 = n_2^{-1}\sum_{i=1}^{n_2} x_{1i}, \qquad (4.36)$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}{n_2 - 1}, \qquad (4.37)$$

$$\bar{x}_2 = n_2^{-1}\sum_{i=1}^{n_2} x_{2i}, \qquad (4.38)$$

and

$$\hat{\bar{\beta}}^2 = (\bar{x}_1 - \bar{x}_2)^2. \qquad (4.39)$$

A *2-dim* output surface can then be constructed using $\left\{ Z_{ASemiP}^{(ij)} \right\}_{i=1,\, j=1}^{R-n-1,\, C-n-1}$, or

$$\mathbf{Z}_{ASemiP} = \begin{bmatrix} Z_{ASemiP}^{(11)}, & \cdots, & Z_{ASemiP}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{ASemiP}^{[(R-n-1)1]}, & \cdots, & Z_{ASemiP}^{[(R-n-1)(C-n-1)]} \end{bmatrix}. \qquad (4.40)$$

4.3.2.2. Asymmetric Variance Test (AVT)

The nonparametric score metric presented in this subsection is arguably the most compact expression for combining-sample based anomaly detection. Transforming $\mathbf{W}_2 = \left(\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2}\right)$ and $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1}\right)$, as described in Section 3.2, to sequences $x_2 = \left(x_{21}, \cdots, \ x_{2n_2}\right)$ and $x_1 = \left(x_{11}, \cdots, \ x_{1n_2}\right)$, we propose to combine these two sequences in the following way in order to determine whether x_1 is anomalous to x_2:

$$Z_{AVT}^{(ij)} = n_2 \frac{\left(S_2^2 - \tau\right)^2}{\hat{\varsigma}_2^2} \tag{4.41}$$

where, $\tau > 0$ a constant chosen by the user, S_2^2 is shown in (4.37), and—for $\bar{x}_2$ shown in (4.38),

$$\hat{\varsigma}_2^2 = \sum_{j=1}^{n_2} \frac{\left[\left(x_{2j} - \bar{x}_2\right)^2 - S_2^2\right]^2}{n_2 - 1}. \tag{4.42}$$

The key in (4.41) is in the choice of τ. Using the sample concatenation shown in (4.31), we choose τ to be

$$\tau = S_{union}^2 = \sum_{i=1}^{n_0} \frac{\left(t_i - \bar{t}\right)^2}{\left(n_0 - 1\right)}, \tag{4.43}$$

where $\bar{t}$ is shown (4.33).

Notice that despite of τ being treated as a constant in (4.41), the value of τ will vary at each testing location (*ij*) in $\mathbf{X}$.

Using (4.41), (4.42), and (4.43), with the data transformation approach described in Section 3.2, constitutes the AVT anomaly detector. High values of (4.41) indicates that $x_2 = \left(x_{21}, \cdots, \ x_{2n_2} \right)$ is anomalous to $x_1 = \left(x_{11}, \cdots, \ x_{1n_2} \right)$.

A *2-dim* output surface can be constructed using $\left\{ Z_{AVT}^{(ij)} \right\}_{i=1, \ j=1}^{R-n-1, \ C-n-1}$, as shown earlier for other detectors, or

$$
\mathbf{Z}_{AVT} = \begin{bmatrix} Z_{AVT}^{(11)}, & \cdots, & Z_{AVT}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{AVT}^{[(R-n-1)1]}, & \cdots, & Z_{AVT}^{[(R-n-1)(C-n-1)]} \end{bmatrix}.
\tag{4.44}
$$

4.3.2.3.   Analysis of Variance (ANOVA)

This subsection briefly presents the ANOVA test statistic—a parametric approach under the normal distribution—and its application to anomaly detection. The ANOVA method [32] attracted our attention because it is one of the most widely used statistical techniques in various fields.

The same two samples (see Subsection 2.2.3) $\mathbf{W}_2 = \left(\mathbf{y}_{21}, \cdots, \ \mathbf{y}_{2n_2} \right)$ and $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right)$ are transformed, as described in Section 3.2, to sequences $x_2 = \left(x_{21}, \cdots, \ x_{2n_2} \right)$ and $x_1 = \left(x_{11}, \cdots, \ x_{1n_2} \right)$. Both sequences are combined using the following formula [32]:

81

$$Z_{ANOVA}^{(ij)} = \frac{\sum_{u=1}^{2} n_u \left(\bar{x}_u - \bar{\bar{x}}\right)^2}{S^2} \qquad (4.45)$$

where, S^2 is defined in (4.34), $\bar{x}_u$ $(u=1,2)$ are defined in (4.36) and (4.38), and

$$\bar{\bar{x}} = \sum_{u=1}^{2} \sum_{h=1}^{n_u} \frac{x_{uh}}{n_u}.$$

High values of (4.45) indicate that x_1 is anomalous to x_2.

This detector's output surface testing $\mathbf{X}$ is constructed as

$$\mathbf{Z}_{ANOVA} = \begin{bmatrix} Z_{ANOVA}^{(11)}, & \cdots, & Z_{ANOVA}^{[1(C-n-1)]} \\ \vdots & \vdots & \vdots \\ Z_{ANOVA}^{[(R-n-1)1]}, & \cdots, & Z_{ANOVA}^{[(R-n-1)(C-n-1)]} \end{bmatrix}. \qquad (4.46)$$

4.4. Summary

Multivariate and univariate techniques and their adaptations to the anomaly detection problem were discussed in this chapter. These techniques will be used for comparison purposes in later chapters. Multivariate techniques are the natural choice for using HS imagery, because this data type is multivariate. Among the various multivariate techniques, the anomaly detection technique, known as the RX anomaly detector, is the most popular in the industry for its utility and, hence, has become the baseline technique for comparison purposes.

This chapter also proposed to use a univariate semiparametric model for anomaly detection applications, which for the best of our knowledge is done for the first time. The motivation for using the semiparametric model is for handling mixtures of HS samples for anomaly detection applications, as discussed in Chapter 1, where anomaly detectors are constantly faced with testing cases in real HS imagery, such that, $g_0 \neq g_1$ but g_0 (from a mixture of two or more material types) significantly overlaps g_1 (from a material type belonging to that mixture). In the context of anomaly detection, it is desirable to relax this difference, because transitions of spectral group regions in $\mathbf{X}$ offer these testing cases and, consequently, degrade performance of existing anomaly detectors testing $\mathbf{X}$. Performance degradation can be clearly observed in Chapter 7 through experiment results testing real HS imagery. Alternative univariate scoring metrics (AsemiP, AVT, and ANOVA) were also proposed. The multivariate detectors (RX, FLD, DPC, and EST) and univariate detectors (SemiP, AVT, AsemiP, and ANOVA) are used for comparison studies in later chapters.

Chapter 5   Power Using Idealized Spectral Samples

5.1. Introduction

Chapter 2 presented, among other topics, the segmentation and characterization of a challenging HS data cube, quantifying in the process an important fact: in analyzing or testing small blocks of data in $\mathbf{X}$, group mixtures (E2, E3, E4, E5, and E6—see Subsection 2.4.1) are expected to exist in real HS imagery, and, as it will be shown in Chapter 6 (TV anomaly detection) and Chapter 7 (GV anomaly detection), if they are ignored, they can significantly degrade the performances of anomaly detectors. Since this fact has been ignored in the HS research community, anomaly detection has been playing a minor role in surveillance or targeting applications. This chapter aims at quantifying the impact spectral mixtures (heterogeneous samples) have on power performances of both classes of anomaly detection approaches (multivariate and univariate), via simulation experiments. These outcomes will be put in the context of what is desired or undesired for anomaly detection applications in order to shed some lights on performance expectations of both classes of detectors on real HS imagery.

The sample structure of $\mathbf{W}_2$ and $\mathbf{W}_1$ will be specified using results from Chapter 2, and these samples will be shared by all the detectors discussed in the Chapter 4. In particular, specific null and alternative hypothesis models will be specified as mixtures of multivariate normal distributions representing idealized spectral samples. For more flexibility, a bias term and a shape term will be added to the estimated spectral mean of G4 (target group) in order to isolate the effect of

changing illumination conditions (bias) from that of different spectral shapes (material spectral distinction).

The remainder of this chapter is organized as follows: Section 5.2 specifies parameters for the null/alternative hypothesis models; Section 5.3 presents the plan used to conduct simulation experiments for this chapter, Section 5.4 discusses simulation results focused on the impact of spectral bias/shape and mixtures on detectors' performances; and Section 5.5 summarizes and concludes this chapter.

5.2. Heterogeneous Models to Study Detection Power

This section describes null/alternative hypothesis models whose samples will be generated and used as inputs for the detectors described in Chapter 4. In particular, the null hypothesis H_0 in (2.11) and the alternative hypothesis H_1 in (2.12) were used to conduct simulation experiments for this chapter. For convenience, we repeat them here with the trial index w, i.e.,

NULL H_0
$$
\begin{aligned}
\mathbf{W}_2^{(w)} &= \left(\mathbf{y}_{21}^{(w)}, \cdots, \ \mathbf{y}_{2n_2}^{(w)}\right) \ \sim i.i.d. \ \ g_2(\mathbf{y}\,|\,\boldsymbol{\theta}) = \sum_{a=1}^{3} g(\mathbf{y}\,|\,\boldsymbol{\theta}_a)\pi_a \\
\mathbf{W}_1^{(w)} &= \left(\mathbf{y}_{11}^{(w)}, \cdots, \ \mathbf{y}_{1n_1}^{(w)}\right) \ \sim \ i.i.d. \ \ g_1(\mathbf{y}\,|\,\boldsymbol{\theta}) = g(\mathbf{y}\,|\,\boldsymbol{\theta}_3)
\end{aligned}
\tag{5.1}
$$

and

ALTERNATIVE H_1
$$
\begin{aligned}
\mathbf{W}_2^{(w)} &= \left(\mathbf{y}_{21}^{(w)}, \cdots, \ \mathbf{y}_{2n_2}^{(w)}\right) \ \sim i.i.d. \ \ g_2(\mathbf{y}\,|\,\boldsymbol{\theta}) = \sum_{a=1}^{3} g(\mathbf{y}\,|\,\boldsymbol{\theta}_a)\pi_a \\
\mathbf{W}_1^{(w)} &= \left(\mathbf{y}_{11}^{(w)}, \cdots, \ \mathbf{y}_{1n_1}^{(w)}\right) \ \sim \ i.i.d. \ \ g_1(\mathbf{y}\,|\,\boldsymbol{\theta}) = \sum_{a=1}^{4} g(\mathbf{y}\,|\,\boldsymbol{\theta}_a)\rho_a
\end{aligned}
\tag{5.2}
$$

where definitions and assumptions are given in (2.11) and (2.12), respectively, and parameter specifications are given in Subsection 2.4.4.

Recall from Subsection 2.4.4 that we arbitrarily fixed proportions $\left(\pi_1, \pi_2, \pi_3\right) = \left(0.4, 0.3, 0.3\right)$ in order to expose the detectors to one of the harder null hypotheses observed in $\mathbf{X}$, i.e., a sample representing event E5 with three groups (G1, G2, G3) versus event E1 with G3. Recall also from Subsection 2.4.4 that the average magnitude of G3 is relatively close to that of G4.

Using (5.2) as the alternative, and systematically varying contribution proportions $\rho_a \left\{a = 1,2,3,4\right\}$, will allow us to check the robustness of these detectors as a function of an alternative that becomes gradually harder relative to the null hypothesis. Certain combinations of $\left(\rho_1, \rho_2, \rho_3, \rho_4\right)$ can determine the level of difficulty imposed on these detectors under H_1 relative to H_0. For instance, H_1 would correspond to H_0 by setting $\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(0.0, 0.0, 1.0, 0.0\right)$ and, thus, by using a small variation of this setting, e.g., $\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(0.0, 0.0, 0.9, 0.1\right)$ with $\hat{\boldsymbol{\mu}}_4 \neq \hat{\boldsymbol{\mu}}_k \ \left(k = 1, \cdots, 3\right)$, H_1 would represent a *hard* alternative hypothesis to detect; alternatively, by setting $\left(\rho_1, \rho_2, \rho_3, \rho_4\right) = \left(0.0, 0.0, 0.0, 1.0\right)$ with $\hat{\boldsymbol{\mu}}_4 \neq \hat{\boldsymbol{\mu}}_k \ \left(k = 1, \cdots, 3\right)$, H_1 would represent an *easy* alternative hypothesis using the same detector.

Since we are also interested in studying the independent effect of spectral magnitude from that of spectral shape on detectors' performances, we added more flexibility to H_1 for this chapter. In particular, we decomposed the mean vector $\boldsymbol{\mu}_4$ (G4) into three parts: a baseline column vector $\boldsymbol{\mu}$ having K dimensions, a bias column (not random) vector Δ_B having K dimensions, and a shape column (not random) vector Δ_S having K dimensions. The 4th-group mean vector (a column vector) then is represented by

$$\mu_4 = \mu + \Delta_B + \Delta_S, \tag{5.3}$$

where $\Delta_B = \begin{bmatrix} \phi \\ \phi \\ \phi \\ \phi \\ \phi \\ \vdots \end{bmatrix}_{(1 \times K)}$; $\phi \geq 0$ (scalar), $\Delta_S = \begin{bmatrix} s \\ -s \\ s \\ -s \\ s \\ \vdots \end{bmatrix}_{(1 \times K)}$; $s \geq 0$ (scalar), and $\mu \in \mathbf{R}^K$.

So, for the simulation experiments presented in this chapter, $\{(\hat{\mu}_a)\}_{a=1}^3$ estimated $\{(\mu_a)\}_{a=1}^3$, as mentioned in Subsection 2.4.4, but, differently here, $\tilde{\mu}_4$ estimates μ_4, where

$$\tilde{\mu}_4 = \hat{\mu}_4 + \Delta_B + \Delta_S. \tag{5.4}$$

Notice that Δ_B changes only the bias of $\tilde{\mu}_4$ and Δ_S changes only the shape of $\tilde{\mu}_4$. This decomposition will allow us to show the difference between *desired* and *undesired* performances of anomaly detectors through the influence of bias and shape changes imposed in $\tilde{\mu}_4$ under H_1 in (5.2).

A simulation plan for this chapter is described next.

5.3. Simulation Plan

Setting the number of bands $K = 120$, idealized pseudo samples will be generated according to models (5.1), (5.2), and (5.4) as follows:

i. Values of $\hat{\boldsymbol{\mu}}_a$ $(a = 1, \cdots, 4)$ and π_a $(a = 1, \cdots, 3)$ under H_0 and H_1 in (5.1) and (5.2), respectively, will be fixed throughout the simulations, but the values of $\tilde{\boldsymbol{\mu}}_4$ in (5.4) will vary according to $(\phi, s) = (19.95, 5.27), (19.95, 0.00), (0.00, 5.27)$, see Section 5.4 for more details; and ρ_a $(a = 1, \cdots, 4)$ under H_1 will vary as specified in (2.13). See Subsection 2.4.4, Section 5.2, and Section 5.4 for specifications and justifications.

ii. Using (i), with fixed sample sizes $n_2 = n_1 = 400$, multivariate random samples $\left(\mathbf{y}_{21}^{(w)}, \cdots, \mathbf{y}_{2n_2}^{(w)} \right)$ and $\left(\mathbf{y}_{11}^{(w)}, \cdots, \mathbf{y}_{1n_1}^{(w)} \right)$ in (5.1) will be generated independently of each other, and, likewise, additional random samples in (5.2) will be generated independently of each other. Random samples generated using (5.1) will be independently generated from those generated using (5.2). Independently of proportions, samples will be generated every trial from multivariate normal distributions $\left\{ N\left(\hat{\boldsymbol{\mu}}_a, \hat{\boldsymbol{\Sigma}}_a \right) \right\}_{a=1}^{3}$ to represent spectral groups G1, G2, and G3, respectively, and $N\left(\tilde{\boldsymbol{\mu}}_4, \hat{\boldsymbol{\Sigma}}_3 \right)$ to represent G4 (recall from Subsection 2.4.4 that, because of confidentiality, we use $\hat{\boldsymbol{\Sigma}}_3$ to represent spectral covariance matrices of G3 and G4). Parameter values for these distributions are specified in Subsection 2.4.4 and Section 5.2. Trials are indexed by w, where $w = 1, \cdots, m$, and the sample sizes n_2 and n_1 are fixed at once to 400.

iii. Attaining Cutoff Thresholds/Estimating Type I Errors: $(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ is set once and for all, and values of $\{\mathbf{W}_2^{(w)}\}_{w=1}^{m}$ and $\{\mathbf{W}_1^{(w)}\}_{w=1}^{m}$ ($m = 1500$) will be generated according to model (5.1) and introduced to the detectors presented in Chapter 4. These detectors will yield *1500* output results each. Each set of *1500* results will be used to estimate an empirical PDF, and a cutoff threshold will be attained by applying the standard quantile method to the empirical PDF. The desired type I error probability (ε) for the quantile method is fixed at once for all the detectors to $\varepsilon = 0.05$. The detectors' corresponding cutoff thresholds will be applied to the corresponding detectors' output results in order to estimate the type I error $(\hat{\varepsilon})$ for each detector. The type I error is estimated by counting the number of trials m_1 that satisfy the detector's output values being greater than the detector's cutoff threshold and computing the ratio $\hat{\varepsilon} = m_1 / m$. (Since the sampling variability involved in estimating the quantiles in this way decreases as a function of *m*, a large *m* [e.g., $m = 1500$] ensure us that these estimates are asymptotically unbiased.)

iv. Estimating the Type I Error: Additional values of $\{\mathbf{W}_2^{(w)}\}_{w=1}^{m}$ and $\{\mathbf{W}_1^{(w)}\}_{w=1}^{m}$ ($m = 1500$) will be generated according to (*ii*) and introduced to both detectors, where the detectors' corresponding cutoff thresholds will be applied to the corresponding detectors' output results in order to estimate the type I error $(\hat{\varepsilon})$ of each detector. The type I error is estimated by counting the number of trials m_1 that satisfy the detector's output values being greater than the detector's

cutoff threshold and computing the ratio $\hat{\varepsilon} = m_1 / m$. (Additional samples will be generated for this step to ensure that the sample size is adequate for the estimation of ε.)

v. Estimating the Power (1.0 – Type II Error): Using $(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ and combinations of $\rho_k \; (k = 1, \cdots, 4)$ and $\tilde{\mu}_4$ from (5.4), additional values of $\{ \mathbf{W}_2^{(w)} \}_{w=1}^m$ and $\{ \mathbf{W}_1^{(w)} \}_{w=1}^m$ (m = 1500) will be generated according to model (5.2) and introduced to all the detectors, where the detectors' corresponding cutoff thresholds will be applied to the corresponding detectors' output results in order to estimate the *power* $(\hat{\theta})$ of each detector for a given set of combinations $\rho_k \; (k = 1, \cdots, 4)$ and $\tilde{\mu}_4$. The type II error is estimated by counting the number m_2 of trials that satisfy the detector's output values being lower then the detector's cutoff threshold and computing the ratio m_2 / m, such that $\hat{\theta} = 1.0 - (m_2 / m)$.

vi. Estimating Performance Confidence Intervals (CI): Variability will be checked via confidence interval estimations for (iv)

$$\bar{\varepsilon} \pm z_{\alpha/2} \sqrt{\frac{1}{U-1} \sum_{u=1}^U (\hat{\varepsilon}_u - \bar{\varepsilon})^2} \quad \text{and} \quad \text{(v)} \quad \bar{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{U-1} \sum_{u=1}^U (\hat{\theta}_u - \bar{\theta})^2}, \quad \text{where}$$

$z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ is the $1 - \alpha/2$ quantile $(\alpha = 0.05)$ of the standard normal distribution, random quantities simulated will be indexed by u for a total number of simulation repetitions U = 2000, $\bar{\varepsilon} = U^{-1} \sum_{u=1}^U \hat{\varepsilon}_u$, and $\bar{\theta} = U^{-1} \sum_{u=1}^U \hat{\theta}_u$.

5.4. Summary of Results

We conducted multiple simulation experiments and organized the exposition of detectors' results in two parts: (*a*) results showing effects of spectral bias/shape on detectors' power performances; and (*b*) results showing effects of a difficult H_0 (involving sample mixtures) on the power of these detectors, as they test different levels of complexity under H_1 (easy, moderate, and hard) relative to H_0.

As it will be shown later in Chapter 7, when properly implemented for real applications, the two-step univariate anomaly detection approach works significantly better suppressing the background clutter in real HS imagery, while accentuating anomalous objects, than does the conventional multivariate anomaly detection approach. For this reason, we used as a criterion for exploring specific parameter combinations under H_1 that of measuring power degradation relative to the calibrated performances of the two-step univariate anomaly detectors and, in the process, record the behavior of multivariate anomaly detectors, as they shared the same input samples used by the two-step univariate detectors.

In order to achieve this goal, we initiated the simulation process by arbitrarily setting some of the parameters, and followed with systematic changes to other parameters, as shown in two subsections within this section, Subsection 5.4.1 and Subsection 5.4.2.

Subsection 5.4.1 discusses the effects of spectral bias/shape differences on these detectors, using the simulation plan outlined in Section 5.3 using a fixed setting for (π_1, π_2, π_3) in (5.1) and (5.2), a fixed setting for $(\rho_1, \rho_2, \rho_3, \rho_4)$ in (5.2), and three combinations of parameter settings for (ϕ, s), which changes $\tilde{\mathbf{\mu}}_4$ in (5.4).

Subsection 5.4.2 discusses the effect of mixture proportions using the same

simulation plan over eight combinations of parameter settings for $\left(\rho_1, \rho_2, \rho_3, \rho_4\right)$ and

seven combinations of parameter settings for (ϕ, s) in $\tilde{\mathbf{\mu}}_4$.

For the results shown in Subsection 5.4.1, it is desired that anomaly detection

approaches are less sensitive to spectral bias but more sensitive to spectral shape,

because high sensitivity to bias differences implies that spectral samples of a

particular material under different illumination conditions may be detected as

anomalous to each other, on the other hand, low sensitivity to spectral shape

differences implies that objects that can blend very well within background clutter

(e.g., camouflaged sniper) may not be detected as scene anomalies, although their

spectral shapes—not their spectral bias—may be significantly different from those of

other materials forming the background clutter.

For results shown in Subsection 5.4.2, it is desired that anomaly detection

approaches are more sensitive to the presence of target samples (G4) in the

corresponding sample mixture under H_1 , given that H_0 also consists of a sample

mixture.

5.4.1. Impact of Spectral Bias/Shape on Detection Performances

Performance results from an experiment that emphasizes the impacts of

spectral bias and spectral shape, independently of each other, on anomaly detectors

are tabulated in Table 5.1 for the multivariate approach and in Table 5.2 for the two-

step univariate approach. The choices of parameter settings labeled as A, B, and C,

see (5.5) below, were particularly chosen to show extreme performance differences of

these detectors, as relatively high bias and/or high shape differences are presented

under H_1. In particular, through preliminary parameter explorations, we learned that, using the *minimum* component value in $\hat{\boldsymbol{\mu}}_4$ as baseline [see (5.4)], a 25% change on bias and a 7% change on shape, were sufficient to cause high impacts on these detectors' performances. Since the minimum component value in $\hat{\boldsymbol{\mu}}_4$ is 79.486 (see Table 5.2), these settings correspond to about $\phi = 19.95$ and $s = 5.56$ in (5.4).

For this simulation experiment, we used three combinations of parameter settings in (ϕ, s)

$$
\begin{array}{ccc}
\text{A} & \text{B} & \text{C} \\
(\phi, s) = (19.95, 5.56) & (19.95, 0.00) & (0.00, 5.56)
\end{array}. \qquad (5.5)
$$

Non-zero parameter values in (5.5) impose significant changes in the magnitude and/or shape of $\tilde{\boldsymbol{\mu}}_4$ [see (5.4)], which in turn cause H_1 in (5.2) to be significantly different from H_0 (5.1) using settings $(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ and $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$. The sample size was fixed to 400, the number of trials m was fixed to 1500, the number of simulation repetitions R was fixed to 2000, and the desired type I error was fixed at $\varepsilon = 0.05$. Using these settings, we estimated the 95% confidence intervals (CI 95%) for the estimated type I error $\hat{\varepsilon}_u$ $(u = 1 \cdots 2000)$ and estimated power $\hat{\theta}_u$ $(u = 1 \cdots 2000)$ for each of these detectors. Since imposed bias/shape changes under H_1 in (5.2) do not interfere with the type I error estimates [see (5.1)], the confidence intervals for type I error were estimated only once per detector, see Table 5.1.

Table 5.1 tabulates results from testing the multivariate anomaly detectors

using these settings, and Table 5.2 tabulates results from testing the two-step

univariate anomaly detectors. All detectors shared the idealized samples.

Table 5.1. Spectral bias/shape impact on power of multivariate detectors (*Det*), where parameter (*Par*) combinations of (ϕ, s) are labeled as A, B, and C, according to the scheme of (5.5).

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$			
		Type I $(\hat{\varepsilon}_u; CI\ 95\%)$		*Power* $(\hat{\theta}_u; CI\ 95\%)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
RX		0.048546493	0.052733506		
	A			1.000000000	1.000000000
	B			1.000000000	1.000000000
	C			0.489456621	0.530792153
KRX		0.048646478	0.052673521		
	A			1.000000000	1.000000000
	B			1.000000000	1.000000000
	C			0.463635719	0.518152153
FLD		0.044824983	0.051990016		
	A			1.000000000	1.000000000
	B			1.000000000	1.000000000
	C			0.460359717	0.502851367
EST		0.047968719	0.057469123		
	A			1.000000000	1.000000000
	B			1.000000000	1.000000000
	C			0.435351767	0.488190723
DPC		0.047510460	0.056507740		
	A			1.000000000	1.000000000
	B			1.000000000	1.000000000
	C			0.441806302	0.509452692

Results in Table 5.1 and Table 5.2 show that detector performances within each table

are comparable among the detectors within these tables, and clearly illustrate the

sensitivity of each approach (multivariate or two-step univariate) to different spectral

features, bias or shape. Multivariate anomaly detectors are sensitive to bias (see, for instance, *Det* RX *Par* B in Table 5.1) and insensitive to shape (see, for instance, RX *Par* C in Table 5.1). On the other hand, the two-step univariate anomaly detectors respond inversely to results shown in Table 5.1 (see, for instance, AsemiP *Par* B and

Table 5.2. Spectral bias/shape impact on power of two-step univariate detectors (*Det*), where parameter (*Par*) combinations of (ϕ, s) are labeled as A, B, and C, according to the scheme of (5.5).

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$			
		Type I $(\hat{\varepsilon}_u; CI\ 95\%)$		*Power* $(\hat{\theta}_u; CI\ 95\%)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
SemiP		0.049689992	0.054310007		
	A			1.000000000	1.000000000
	B			0.493442024	0.534620953
	C			1.000000000	1.000000000
AsemiP		0.048396724	0.053763275		
	A			1.000000000	1.000000000
	B			0.497496015	0.537630861
	C			1.000000000	1.000000000
AVT		0.048724038	0.053715961		
	A			1.000000000	1.000000000
	B			0.491576768	0.536911794
	C			1.000000000	1.000000000
ANOVA		0.047864640	0.052557102		
	A			1.000000000	1.000000000
	B			0.479205674	0.527917876
	C			1.000000000	1.000000000

Par C in Table 5.2). Decomposing these two spectral features made it possible to review sensitivity differences between the two classes of detectors; otherwise, perfect

performances would have been shown for both classes of detectors, see, for instance,

KRX *Par* A in Table 5.1 and AVT *Par* A in Table 5.2. We attribute these differences

in sensitivity to the data transformation step in the two-step univariate approach. The

implication of these differences, however, may be critical for applications using real

HS imagery, as it will be discussed shortly.

Starting with two spectral means, $\tilde{\boldsymbol{\mu}}_4[(\phi, s) = (0.00,0.00)]$ and $\hat{\boldsymbol{\mu}}_3$, that are not

too distinct from each other, and using mixture proportion $\rho_4 = 1.0$ and parameter

combination A, significant distinction between H_0 in (5.1) and H_1 in (5.2) exist and,

as shown in Table 5.1 and Table 5.2, both classes of detectors respond with perfect

power. This experiment using parameter combination B simulates testing cases

involving samples of a particular material under different illumination conditions by

removing the average shape difference between i.i.d. random samples using

$g(\mathbf{y} \mid \boldsymbol{\theta}_3) = N(\hat{\boldsymbol{\mu}}_3, \hat{\boldsymbol{\Sigma}}_3)$ and i.i.d. random samples using $g(\mathbf{y} \mid \boldsymbol{\theta}_4) = N(\tilde{\boldsymbol{\mu}}_4, \hat{\boldsymbol{\Sigma}}_3)$, see (5.1),

(5.2) and (5.4), while preserving a significant bias difference between both samples.

Since the data transformation step discussed in Chapter 3 is proposed to remove the

impact of bias and preserve shape differences, the two-step univariate anomaly

detectors could not distinguish between sample sets under H_0 in (5.1) and H_1 in

(5.2), thus, their confidence intervals for power are comparable to their confidence

intervals for type I errors, see Table 5.2. For real anomaly detection applications,

however, this performance behavior is in fact desired, because testing cases involving

samples of a particular material having a distinct spectral shape but under different

illumination conditions are significantly more common in the real world than samples

of materials having similar spectral shapes under a particular illumination condition. (Changes on illumination environment influence the bias, or magnitude, in spectra.)

On the contrary, conducting this experiment with parameter combination C removes the bias, while preserving a significant average shape difference between random samples from $g(\mathbf{y}\,|\,\boldsymbol{\theta}_3) = N\!\left(\hat{\boldsymbol{\mu}}_3, \hat{\boldsymbol{\Sigma}}_3\right)$ and $g(\mathbf{y}\,|\,\boldsymbol{\theta}_4) = N\!\left(\tilde{\boldsymbol{\mu}}_4, \hat{\boldsymbol{\Sigma}}_3\right)$. Results in Table 5.2 show that the two-step univariate detectors do in fact emphasize on shape differences among multivariate samples, while multivariate detectors do not, see Table 5.1 for parameter combination C. The advantage of having sensitivity to spectral shape is that many categories of the difficult problem detecting camouflaged targets falls under the case mimicked by this simulation experiment using parameter combination C, i.e., testing samples of different materials having different spectral shapes, but similar spectral bias, under any illumination condition. For instance, materials composing sniper camouflage suits provide an average spectral magnitude that is comparable to the average spectral magnitude of natural canopy, but most of these camouflage suits have spectral shapes that are different from natural canopy in other regions of the spectrum outside the visible region (e.g., near infrared), these findings were reported in [33].

In summary, it is desired to have an anomaly detection approach that is sensitive to spectral shape and insensitive to spectral bias. These features, however, are not sufficient to address another important issue: detection approach's inability to deal with mixtures of spectral samples in HS imagery. A suitable detection approach for samples involving mixtures would be able to maintain cutoff thresholds relatively low under the challenging null hypothesis in (5.1), hence, becoming potentially more

sensitive to the presence of target samples in the alternative hypothesis in (5.2). We will address next this particular issue.

5.4.2. Impact of Spectral Mixtures on Detection Performances

This subsection shows the benefit of using an anomaly detection approach capable of maintaining a relatively low cutoff threshold under a challenging null hypothesis in (5.1), as it tests the alternative hypothesis in (5.2) for different parameter combinations. This assessment can be made by recording the estimated power of these detectors, as a function of decreasing proportion of G4 (ρ_4) in (5.2) from unity to smaller values. For this simulation experiment, a desired detection technique is expected to yield a superior power relative to performances of alternative approaches.

As the two anomaly detection approaches (multivariate and two-step univariate) are sensitive to different types of spectral features (bias or shape), we will ignore in this part of the simulation whether the detectors' outcomes for the alternative hypothesis are desired or undesired, and calibrate instead their performances to a baseline, i.e., using the simulation plan described in Section 5.3, we will find a combination of parameters that will cause all the detectors to perform about the same, and name it: *calibrated performances*. We explored various combinations of parameter values $(\rho_1, \rho_2, \rho_3, \rho_4, \phi, s)$ and attained calibrated performances using parameter combination $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$ and seven labeled combinations of (ϕ, s):

$$
\begin{array}{ccc}
\text{A1} & \text{A2} & \text{A3} \\
(\phi, s) = (19.95, 5.56) & (17.68, 4.23) & (15.19, 3.60)
\end{array}
$$

$$
, \qquad (5.6)
$$

$$
\begin{array}{cccc}
\text{A4} & \text{A5} & \text{A6} & \text{A7} \\
(\phi, s) = (11.87, 2.39) & (8.34, 1.58) & (5.60, 0.97) & (0.00, 0.00)
\end{array}
$$

where, for the given parameter combination $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$, the label combination A1 presents the *easiest* alternative hypothesis for both detection approaches to detect, and A7 presents the *hardest* alternative hypothesis for both approaches.

Detection performances using the simulation plan outlined in Section 5.3 and parameter combinations A1 through A7 in (5.6) for $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$ led to calibrated results, we created additional challenges to these detectors by changing mixture proportions ρ_k $(k = 1, \cdots, 4)$ under H_1 in (5.2), as mixture proportions π_k $(k = 1, \cdots, 3)$ under H_0 and H_1 in (5.1) and (5.2), respectively, were kept fixed. In particular, π_k $(k = 1, \cdots, 3)$ were arbitrarily set to be approximately equal to each other, and ρ_4 was gradually decreased, as ρ_k $(k = 1, \cdots, 3)$ were set approximately equal to each other satisfying the requirement $\sum_{k=1}^{4} \rho_k = 1.0$. (Performance results using different parameter combinations were not included in this subsection because they were either redundant or did not show significant differences from the results included herein.) Summarized calibrated performances are shown in Table 5.3 and Table 5.4 for multivariate detectors and two-step univariate detectors, respectively, where the type I error CI estimations are

99

shown only once per detector. Table 5.5 through Table 5.10 show detectors' performances using the same simulation plan and seven parameter combinations of $(\rho_1, \rho_2, \rho_3, \rho_4)$, over the seven labeled combinations according to scheme in (5.6). Results within each table from Table 5.3 thru Table 5.10 were achieved for a fixed parameter combination of $(\rho_1, \rho_2, \rho_3, \rho_4)$ over the same seven labeled combinations of (ϕ, s). Actual differences between results shown in Table 5.3 and Table 5.4 for a given parameter combination are unimportant, but the fact that these results gradually decrease using combinations A1 through A7. The difficulty level of the alternative hypothesis in (5.2) was increased by changing the values of $(\rho_1, \rho_2, \rho_3, \rho_4)$, as follows:

Multivariate Detectors (Table 5.5): $\begin{aligned} (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.02, 0.02, 0.01, 0.95) \\ (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.04, 0.03, 0.03, 0.90) \end{aligned}$,

Univariate Detectors (Table 5.6): $\begin{aligned} (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.02, 0.02, 0.01, 0.95) \\ (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.04, 0.03, 0.03, 0.90) \end{aligned}$,

Multivariate Detectors (Table 5.7): $\begin{aligned} (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.07, 0.07, 0.06, 0.80) \\ (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.14, 0.13, 0.13, 0.60) \end{aligned}$,

Univariate Detectors (Table 5.8): $\begin{aligned} (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.07, 0.07, 0.06, 0.80) \\ (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.14, 0.13, 0.13, 0.60) \end{aligned}$,

Univariate Detectors (Table 5.9): $\begin{aligned} (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.20, 0.20, 0.20, 0.40) \\ (\rho_1, \rho_2, \rho_3, \rho_4) &= (0.27, 0.27, 0.26, 0.20) \end{aligned}$,

Univariate Detectors (Table 5.10): $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.32, 0.32, 0.31, 0.05)$.

Tables were organized, such that, for a fixed parameter combination of (ϕ, s), e.g., A4, the alternative hypothesis ranges from the *easiest* (Table 5.3 and Table 5.4)

Table 5.3. Multivariate Detectors' Calibrated Performances

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$			
		Type I $(\hat{\varepsilon}_u; CI\ 95\%)$		*Power* $(\hat{\theta}_u; CI\ 95\%)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
RX	A1	0.047847406	0.052466579	1.000000000	1.000000000
	A2			0.951231671	0.980848328
	A3			0.841845730	0.887154269
	A4			0.716048622	0.770871326
	A5			0.730833174	0.789046825
	A6			0.670805233	0.742074766
	A7			0.485570856	0.531229143
KRX	A1	0.047019216	0.051427223	1.000000000	1.000000000
	A2			0.949160342	0.992473218
	A3			0.861351716	0.901697124
	A4			0.726733423	0.752097634
	A5			0.684823760	0.736574879
	A6			0.637513465	0.680138637
	A7			0.469282102	0.513426218
FLD	A1	0.049634992	0.051939022	1.000000000	1.000000000
	A2			0.971265672	0.999654325
	A3			0.863428543	0.900137528
	A4			0.741682689	0.786720973
	A5			0.673028745	0.699236734
	A6			0.622723046	0.679211037
	A7			0.458174234	0.499305291
EST	A1	0.046972706	0.051540783	1.000000000	1.000000000
	A2			0.932634527	0.957292632
	A3			0.896146341	0.942467833
	A4			0.751691624	0.796222047
	A5			0.703926342	0.760759260
	A6			0.646196725	0.660263981
	A7			0.432609245	0.482932473
DPC	A1	0.047973420	0.054310907	1.000000000	1.000000000
	A2			0.926298163	0.971450998
	A3			0.826298739	0.887845417
	A4			0.754562215	0.804520865
	A5			0.680272746	0.750684321
	A6			0.610924632	0.672084098
	A7			0.443698265	0.505234092

Table 5.4. Univariate Detectors' Calibrated Performances

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.0, 0.0, 0.0, 1.0)$			
		Type I $\left(\hat{\varepsilon}_u; CI\ 95\%\right)$		*Power* $\left(\hat{\theta}_u; CI\ 95\%\right)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
SemiP	A1	0.049315748	0.053449251	1.000000000	1.000000000
	A2			0.929022544	0.978537455
	A3			0.864110188	0.920069811
	A4			0.740582052	0.799857947
	A5			0.716606531	0.760793468
	A6			0.681982684	0.736870315
	A7			0.496821036	0.535907963
AsemiP	A1	0.047081616	0.052158383	1.000000000	1.000000000
	A2			0.931021793	0.986394343
	A3			0.870843425	0.930429148
	A4			0.747906238	0.793700263
	A5			0.723217450	0.772359961
	A6			0.688340311	0.729983092
	A7			0.494286565	0.538417539
AVT	A1	0.047855700	0.052984299	1.000000000	1.000000000
	A2			0.939854540	0.982983456
	A3			0.867436715	0.916278640
	A4			0.740923484	0.793429346
	A5			0.708943277	0.759542720
	A6			0.681309785	0.730628767
	A7			0.492561163	0.537230698
ANOVA	A1	0.050133742	0.054604205	1.000000000	1.000000000
	A2			0.904538671	0.958146032
	A3			0.832734752	0.860029434
	A4			0.713459863	0.769620832
	A5			0.675814342	0.728826260
	A6			0.659110586	0.706203245
	A7			0.479681125	0.525821138

Table 5.5. Multivariate Detection Performances Using Target Contributions

$$\rho_4 = 0.95 \text{ and } \rho_4 = 0.90.$$

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.02, 0.02, 0.01, 0.95)$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.04, 0.03, 0.03, 0.90)$	
		Power $(\hat{\varepsilon}_u ; CI\ 95\%)$		*Power* $(\hat{\theta}_u ; CI\ 95\%)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
RX	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.825180772	0.903339227	0.825180772	0.903339227
	A3	0.555990555	0.621049444	0.555990555	0.621049444
	A4	0.463348966	0.490131033	0.463348966	0.490131033
	A5	0.396139002	0.431491347	0.396139002	0.431491347
	A6	0.334816120	0.369583879	0.334816120	0.369583879
	A7	0.301682679	0.343297320	0.201682679	0.243297320
KRX	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.795328748	0.912645739	0.795328748	0.912645739
	A3	0.484275752	0.536754864	0.484275752	0.536754864
	A4	0.385674314	0.420465197	0.385674314	0.420465197
	A5	0.366737011	0.409743051	0.366737011	0.409743051
	A6	0.321346767	0.367962340	0.321346767	0.367962340
	A7	0.328964844	0.339641464	0.228964844	0.239641464
FLD	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.847091342	0.884873454	0.847091342	0.884873454
	A3	0.490452846	0.534345460	0.490452846	0.534345460
	A4	0.423095088	0.465643454	0.423095088	0.465643454
	A5	0.383078521	0.430396405	0.383078521	0.430396405
	A6	0.358779037	0.386531785	0.358779037	0.386531785
	A7	0.309156374	0.337648751	0.209156374	0.237648751
EST	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.782600875	0.892773145	0.782600875	0.892773145
	A3	0.475666741	0.526485023	0.475666741	0.526485023
	A4	0.443782624	0.499761458	0.443782624	0.499761458
	A5	0.382297420	0.429237810	0.382297420	0.429237810
	A6	0.330616853	0.386435256	0.330616853	0.386435256
	A7	0.308635328	0.356781011	0.208635328	0.256781011
DPC	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.802842413	0.872417029	0.802842413	0.872417029
	A3	0.468507363	0.518149026	0.468507363	0.518149026
	A4	0.439726230	0.483719095	0.439726230	0.483719095
	A5	0.369147613	0.393171322	0.369147613	0.393171322
	A6	0.339664029	0.385387824	0.339664029	0.385387824
	A7	0.317153171	0.358695472	0.217153171	0.258695472

Table 5.6. Univariate Detection Performances Using Target Contributions

$$\rho_4 = 0.95 \text{ and } \rho_4 = 0.90.$$

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1,\cdots,U$ $\varepsilon = 0.05$		$(\pi_1,\pi_2,\pi_3) = (0.4,0.3,0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.02,0.02,0.01,0.95)$		$(\pi_1,\pi_2,\pi_3) = (0.4,0.3,0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.04,0.03,0.03,0.90)$	
		Power $(\hat{\varepsilon}_u; CI\ 95\%)$		Power $(\hat{\theta}_u; CI\ 95\%)$	
Det	Par	Lower Bound	Upper Bound	Lower Bound	Upper Bound
SemiP	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.909659154	0.952889404	0.874303299	0.911885592
	A3	0.785618764	0.833496200	0.775690447	0.849680672
	A4	0.721621835	0.779012731	0.702376401	0.756155405
	A5	0.631820171	0.685204027	0.631657641	0.671817153
	A6	0.586345942	0.625283317	0.585082625	0.643055772
	A7	0.435402217	0.479028743	0.431993225	0.490968901
AsemiP	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.915603467	0.958082184	0.873649822	0.899464574
	A3	0.840036175	0.891523673	0.780914594	0.836971818
	A4	0.713634931	0.769364527	0.713661081	0.761853164
	A5	0.660860339	0.695026209	0.605455535	0.658781068
	A6	0.619268476	0.660771130	0.532986941	0.578476305
	A7	0.453980184	0.514069935	0.423245019	0.484886846
AVT	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.933709213	0.987454674	0.888579268	0.933080954
	A3	0.859801046	0.903526815	0.803244410	0.867227484
	A4	0.746831065	0.798460357	0.724969739	0.781469620
	A5	0.688092370	0.739062780	0.636852430	0.692895861
	A6	0.629869185	0.668237452	0.519073045	0.580765283
	A7	0.489722509	0.539620926	0.446294217	0.495503491
Anova	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.882020533	0.937508645	0.864772451	0.912585297
	A3	0.759627628	0.801957892	0.756760085	0.813233072
	A4	0.689158367	0.739592125	0.719977948	0.761886526
	A5	0.611093528	0.662493658	0.571830266	0.648581003
	A6	0.539474725	0.579035200	0.454181254	0.499717765
	A7	0.447016420	0.488402435	0.385887644	0.438742646

to the *hardest* (Table 5.10), where results using parameter combination A1 shown in Table 5.3 and Table 5.4 shows results for the absolute easiest alternative hypothesis in this simulation and results using A7 in Table 5.10 represent the absolute hardest alternative hypothesis.

Table 5.5 through Table 5.8 show that anomaly detection power performance of all the detectors gradually decline from their calibrated performances (Table 5.3 and Table 5.4), with dramatic degradation experienced by multivariate detectors somewhere between settings $\rho_4 = 0.80$ and $\rho_4 = 0.60$, see Table 5.8. Decreasing the contribution of G4 (target group) in the alternative hypothesis from 1.0 to 0.6 was sufficient to cause this dramatic degradation of these multivariate detectors. The univariate detectors show a slower degradation in those tables; in fact, a more tolerable degradation. As we have been discussing throughout this dissertation, tolerance to mixture is a desired capability for effective anomaly detection, as it will be shown in Chapter 6 and Chapter 7.

We attribute the favorable performances shown in Table 5.6 and Table 5.8 of univariate detectors to their sample combining strategy built into their scoring metrics. The data transformation step of these univariate detectors may play a role in producing favorable detection performances, but we could not isolate that role from the sample combining strategy.

Table 5.7. Multivariate Detection Performances Using Target Contributions

$$\rho_4 = 0.80 \textbf{ and } \rho_4 = 0.60.$$

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1,\cdots,U$ $\varepsilon = 0.05$		$(\pi_1,\pi_2,\pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.07, 0.07, 0.06, 0.80)$ Power $\left(\hat{\theta}_u ; CI\ 95\%\right)$		$(\pi_1,\pi_2,\pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.14, 0.13, 0.13, 0.60)$ Power $\left(\hat{\theta}_u ; CI\ 95\%\right)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
RX	A1	0.733200540	0.782074667	0.000000000	0.000000000
	A2	0.553764986	0.605471812	0.000000000	0.000000000
	A3	0.363348966	0.391049444	0.000000000	0.000000000
	A4	0.196139002	0.220131033	0.000000000	0.000000000
	A5	0.032816120	0.079814913	0.000000000	0.000000000
	A6	0.000000000	0.000000000	0.000000000	0.000000000
	A7	0.000000000	0.000000000	0.000000000	0.000000000
	A1	0.731571219	0.777193378	0.000000000	0.000000000
KRX	A2	0.552849573	0.592645739	0.000000000	0.000000000
	A3	0.360089609	0.416754864	0.000000000	0.000000000
	A4	0.198674672	0.240465197	0.000000000	0.000000000
	A5	0.031751518	0.089743051	0.000000000	0.000000000
	A6	0.000000000	0.000000000	0.000000000	0.000000000
	A7	0.000000000	0.000000000	0.000000000	0.000000000
	A1	0.739562228	0.761744099	0.000000000	0.000000000
FLD	A2	0.555759078	0.589873454	0.000000000	0.000000000
	A3	0.361961217	0.404345460	0.000000000	0.000000000
	A4	0.194326639	0.245643454	0.000000000	0.000000000
	A5	0.038956255	0.084396405	0.000000000	0.000000000
	A6	0.000000000	0.000000000	0.000000000	0.000000000
	A7	0.000000000	0.000000000	0.000000000	0.000000000
	A1	0.725615296	0.764833090	0.000000000	0.000000000
EST	A2	0.583040725	0.622773145	0.000000000	0.000000000
	A3	0.379437886	0.426485023	0.000000000	0.000000000
	A4	0.071523457	0.109761458	0.000000000	0.000000000
	A5	0.000000000	0.000000000	0.000000000	0.000000000
	A6	0.000000000	0.000000000	0.000000000	0.000000000
	A7	0.000000000	0.000000000	0.000000000	0.000000000
	A1	0.728861565	0.769659348	0.000000000	0.000000000
DPC	A2	0.578396968	0.612417029	0.000000000	0.000000000
	A3	0.348908983	0.398149026	0.000000000	0.000000000
	A4	0.079621643	0.123719095	0.000000000	0.000000000
	A5	0.000000000	0.000000000	0.000000000	0.000000000
	A6	0.000000000	0.000000000	0.000000000	0.000000000
	A7	0.000000000	0.000000000	0.000000000	0.000000000

Symbol "#" in a table indicates that convergence was not achieved using the

parameter initialization for the SemiP anomaly detector, see implementation in [31].

Table 5.8. Univariate Detection Performances Using Target Contributions $\rho_4 = 0.80$ and $\rho_4 = 0.60$. Symbol "#" indicates convergence was not achieved using initial parameters.

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1,\cdots,U$ $\varepsilon = 0.05$		$(\pi_1,\pi_2,\pi_3) = (0.4,0.3,0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.07,0.07,0.06,0.80)$ *Power* $\left(\hat{\theta}_u; CI\ 95\%\right)$		$(\pi_1,\pi_2,\pi_3) = (0.4,0.3,0.3)$ $(\rho_1,\rho_2,\rho_3,\rho_4) = (0.14,0.13,0.13,0.60)$ *Power* $\left(\hat{\theta}_u; CI\ 95\%\right)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
SemiP	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.847330572	0.880966261	0.813693839	0.851066836
	A3	0.760797702	0.804269901	0.735067470	0.779696583
	A4	0.712595065	0.766212853	0.683828381	0.738584439
	A5	#	#	#	#
	A6	#	#	#	#
	A7	#	#	#	#
AsemiP	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.847674469	0.887891337	0.829083671	0.874565704
	A3	0.758994682	0.794583920	0.731571435	0.782310240
	A4	0.712890934	0.754461533	0.677066212	0.727534560
	A5	0.543391191	0.590573722	0.533539469	0.589997718
	A6	0.411218079	0.461773307	0.399786923	0.440429952
	A7	0.369442180	0.408530734	0.324301102	0.372755923
AVT	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.841896321	0.872358308	0.838394976	0.889014590
	A3	0.759896179	0.796683731	0.726182778	0.777605066
	A4	0.711378531	0.757163400	0.677657743	0.728241211
	A5	0.549103918	0.603195415	0.513841817	0.561613723
	A6	0.417853942	0.439021960	0.414720998	0.458651897
	A7	0.372508762	0.422576169	0.329397866	0.370103762
ANOVA	A1	1.000000000	1.000000000	1.000000000	1.000000000
	A2	0.843689645	0.885250105	0.818258933	0.853094950
	A3	0.751328709	0.804916698	0.740077156	0.790668572
	A4	0.681374569	0.739190354	0.672524142	0.768465956
	A5	0.548981638	0.598102132	0.543996648	0.592347415
	A6	0.410318052	0.468235759	0.385480253	0.430886928
	A7	0.365159282	0.418096914	0.311386414	0.360965917

The symbol "#" is shown in Table 5.8 through Table 5.10.

We attribute the dramatic power degradation of the multivariate detectors shown in Table 5.7 to their inabilities to yield cutoff thresholds that are relatively low under H_0 in (5.1) with respect to these detectors' responses under H_1 in (5.2), using the parameter combinations in (5.6), and G4 proportion combinations $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.07, 0.07, 0.06, 0.80)$ and

$(\rho_1, \rho_2, \rho_3, \rho_4) = (0.14, 0.13, 0.13, 0.60)$. The consequence of this inability will be better appreciated in Chapter 6, where simulated multispectral cubes will be tested using all these detectors in the context of a top-view anomaly detection application, and in Chapter 7, where real HS data cubes will be tested in an actual ground to ground anomaly detection application.

Table 5.9 through Table 5.10 tabulates additional power degradation of two univariate detectors (SemiP and AsemiP), as their responses under H_1 in (5.2) overlaps significantly more their responses under H_0 in (5.1), especially in Table 5.10 and for parameter combinations A6 and A7 in Table 5.9. (Performances testing AVT and ANOVA using parameter combinations shown in Table 5.9 and Table 5.10 were comparable to those of AsemiP, thus, we tabulated only AsemiP's results in those tables.) In Table 5.9, however, under a high shape difference (A1), these detectors could still hold a power level corresponding to their calibrated performances in Table 5.4, including for parameter combination $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.27, 0.27, 0.26, 0.20)$. Holding power performances as shown

illustrates these detectors' low sensitivity under H_0 in (5.1) and relatively high

sensitivity to an imposed 20% contribution of G4 to the mixture under H_1 in (5.2).

Multivariate detection performances using $\rho_4 < 0.60$ were unchanged from

their results using $\rho_4 = 0.60$.

Table 5.9. Univariate Detection Performances Using Target Contributions $\rho_4 = 0.40$ and $\rho_4 = 0.20$. (AVT and ANOVA detectors performed comparably with AsemiP detector, thus, their results are not included in this table or in Table 5.10.)

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.20, 0.20, 0.20, 0.40)$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.27, 0.27, 0.26, 0.2$	
		Power $\left(\hat{\theta}_u; CI\ 95\%\right)$		Power $\left(\hat{\theta}_u; CI\ 95\%\right)$	
Det	*Par*	*Lower Bound*	*Upper Bound*	*Lower Bound*	*Upper Bound*
SemiP	A1	0.985729013	1.000000000	0.955967351	0.990179398
	A2	0.813993425	0.851844302	0.722426477	0.762676989
	A3- A7	#	#	#	#
AsemiP	A1	1.000000000	1.000000000	0.928694491	0.970056224
	A2	0.800279305	0.842764417	0.734767327	0.788677427
	A3	0.714281832	0.761782503	0.560643222	0.614028643
	A4	0.633847121	0.687343948	0.473282337	0.531158130
	A5	0.490957338	0.548122835	0.334160903	0.385051201
	A6	0.378573463	0.435264269	0.197696974	0.246697928
	A7	0.289119305	0.338127854	0.113934239	0.171447510

Table 5.10. Univariate Detection Performances Using Target Contribution $\rho_4 = 0.05$.

$n_2 = n_1 = 400$ $m = 1500$ $U = 2000$ $u = 1, \cdots, U$ $\varepsilon = 0.05$		$(\pi_1, \pi_2, \pi_3) = (0.4, 0.3, 0.3)$ $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.32, 0.32, 0.31, 0.05)$	
		Power $\left(\hat{\theta}_u; CI\ 95\% \right)$	
Det	*Par*	*Lower Bound*	*Upper Bound*
SemiP	A1	0.129408121	0.181614463
	A2	0.048106337	0.054150774
	A3-	#	#
AsemiP	A7	0.122216924	0.179680262
	A1	0.049159128	0.053750184
	A2	0.012320469	0.018847234
	A3	0.000000000	0.000000000
	A4	0.000000000	0.000000000
	A5	0.000000000	0.000000000
	A6	0.000000000	0.000000000
	A7		

In essence, results tabulated in Table 5.5 through Table 5.14 illustrate the robustness of these two anomaly detection approaches, as a function of increased complexity under H_1. In this simulation, the complexity level under H_1 could be significantly increased by gradually decreasing ρ_4 and adjusting ρ_1, ρ_2, and ρ_3, accordingly. By following this procedure, we could assess detectors' abilities to maintain a rather low cutoff threshold, which in turn yield higher power under challenging null and alternative hypotheses. Again, it is difficult for us to separate with absolute certainty how much of the univariate detectors' favorable performances are due to their data transformation step or to their scoring metric step.

5.5. Summary and Conclusions

This chapter provided simulation experiment results that show relative strengths between two anomaly detection approaches on training and testing hypotheses involving spectral mixtures. Having all nine detectors trained with sample mixtures under H_0 in (5.1), we examined the impact on their power performances for detecting the presence of G4 (target group) by varying key parameters under H_1 in (5.2). Through this study, we were able to separate the impact on detection performances owing to three factors considered important in this dissertation: spectral bias, spectral shape, and varying mixture proportions of G4 under H_1. Under these factors, the two-step univariate approach performed better and more desirably relative to the multivariate approach, as shown in tabulated results in this chapter. Those favorable results by the two-step approach have some key implications: (i) this approach is more likely not to detect HS samples of a particular material under different illumination conditions as being anomalous to each other (low sensitivity to spectral bias); (ii) this approach is more likely to detect the presence of a particular material that can blend very well with the background scene, although this material may have an average spectral shape that is distinct from those of background spectra (high sensitivity to spectral shape); and (iii) this approach is more likely to yield a rather low cutoff threshold under difficult null hypotheses (involving sample mixtures), which in turn can produce higher power under similarly challenging alternative hypotheses, where a target material may be only partially represented in the test sample (low sensitivity to mixtures in null hypotheses). The three

implications can play a major role in achieving desired performances in real anomaly

detection applications, as it will be shown in Chapter 7 on real HS imagery. Chapter 6

implements both anomaly detection approaches to a TV anomaly detection

application in order to examine (iii) using simulated multivariate data cubes.

Chapter 6   Power Using Idealized Top View Cubes

6.1. Introduction

We seek to demonstrate in this chapter the differences in performances among different anomaly detection techniques using idealized data cubes for the case of controlled top-view background configuration scenarios.

Different from the simulation experiments discussed in Chapter 5, the simulation experiment discussed in this chapter uses idealized pseudo cubes representing smaller versions of real HS data cubes. This chapter's main goal is to assess the power of detection approaches discussed in Chapter 4 for detecting anomalies using these idealized pseudo cubes, and to use the same standard statistical method discussed in Chapter 5 to assess power estimates, at a fixed sample size but varying type I error. To achieve this goal, we start by generating a large number of idealized *training* cubes using three different background clutter configurations, and test these cubes with various anomaly detectors so that cutoff thresholds and type I errors can be estimated from these detectors' output surfaces. (Each pixel in a given output surface corresponds to a trial outcome using the corresponding detector.) We then generate another large set of idealized *test* cubes using the same three background configurations and added *targets* and test these cubes using the same detectors and their corresponding cutoff thresholds so that their powers can be estimated from their output surfaces. Tabulated results and ROC curves will be shown for this simulation experiment.

A conjecture: the anomaly detection algorithms that are relatively insensitive to increasingly more complex background configurations should produce higher power for detecting embedded targets in test cubes. This desired result will not be as obvious for a homogeneous clutter background having simple spatial configuration, but more obvious for a heterogeneous background having a more complex spatial configuration relative to target scales in the imagery.

Since this simulation experiment was designed to estimate detectors' power over large numbers of experiment trials and repetitions, all the detectors discussed in Chapter 4 were used in this simulation experiment, excluding kernel RX, or KRX, owing to its prohibitive amount of computational time required to test a single cube.

6.2. Notations and Definitions

Many of the notations in this section apply only to this chapter, but an effort was made to match key notations already made in previous chapters, as appropriate. Also, since the mechanics of testing samples at different window locations in a single data cube represent different experimental trials and experimental repetitions require independent constructions of data cubes, for computational reasons, we settled for generating idealized data cubes having only *5* bands in order to mimic multispectral (MS) data cubes rather than HS data cubes (cubes greater than 100 bands). And, since we do not have access to actual MS data cubes, the parameter settings used to generate idealized samples for cube construction were not based on the physics of spectroscopy, but were chosen to generate samples of different *classes* distinct enough from each other in order to illustrate the differences in performance among detectors, as a function of varying spatial complexity in background configurations.

114

(A multivariate sample representing a class in this simulation experiment means a sample generated by setting parameters to particular values in a corresponding PDF, in this case, a normal PDF.)

Two sets of idealized MS cubes were generated: training $\mathbf{B}_h^{(g)}$ (null hypothesis cubes; sometimes also referred to as $B_h^{(g)}$ in this chapter) and testing $\mathbf{BT}_z^{(g)}$ (alternative hypothesis cubes; sometimes also referred to as $BTz^{(g)}$ in this chapter), where these cubes have the format of $\mathbf{X}$ in (2.4) with $R = C = 256$ and $K = 5$; $h = 1, \cdots, 3$ indexes three different kinds of training background configurations; $z = 1, \cdots, 4$ indexes four different kinds of target-background configurations; and $g = 1, \cdots, 1500$ indexes independent experimental repetitions.

A random sample representing a single background class in $\mathbf{B}_h^{(g)}$, or $\mathbf{BT}_z^{(g)}$, are i.i.d. normal, or

$$\left\{ C_k^{(g,r,c)} \right\}_{r=i,c=j}^{I,J} \sim \ i.i.d. \ \ N\left(\mu_k, \Sigma\right), \tag{6.1}$$

where, $C_k^{(g,r,c)}$ and μ_k are column vectors $\left(C_k^{(g,r,c)} \in \mathbf{R}^K; \mu_k \in \mathbf{R}^K \right)$; $\Sigma \in \mathbf{R}^{K \times K}$; $k = 1, \cdots, 6$ indexes a background class; and $r = i, c = j$ $\left(1 \leq i, \cdots, I \leq R; 1 \leq j, \cdots, J \leq C \right)$ index the row and column, respectively, in $\mathbf{B}_h^{(g)}$. (Note that, since Σ is shared by all background classes, μ_k determines the kth background class, where $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6$. Parameter specifications are presented in Subsection 6.3.2)

Using (6.1) and parameter specifications from Subsection 6.3.2, a training cube $\mathbf{B}_h^{(g)}$ can be constructed to spatially represent different background class regions; for example, for a particular background configuration (see Figure 6.1 below),

$$\mathbf{B}_2^{(g)} = \left[\mathbf{C}_1\, \mathbf{C}_2\, \mathbf{C}_3\, \mathbf{C}_4\, \mathbf{C}_5\, \mathbf{C}_6\right]^{(g)},$$ (6.2)

where, $\mathbf{C}_1 = \left[C_1^{(g,r,c)} : r = 1, \cdots, R; c = 1, \cdots, J_1 < C\right]$,

$\mathbf{C}_2 = \left[C_2^{(g,r,c)} : r = 1, \cdots, R; c = J_1 + 1, \cdots, J_2 < C\right]$,

$\mathbf{C}_3 = \left[C_3^{(g,r,c)} : r = 1, \cdots, R; c = J_2 + 1, \cdots, J_3 < C\right]$,

$\cdots, \mathbf{C}_6 = \left[C_6^{(g,r,c)} : r = 1, \cdots, R; c = J_5 + 1, \cdots, C\right]$.

Fig. 6.1 shows one of the background configurations $\mathbf{B}_2$ (background

Training Cube B_2

Figure 6.1. A training (null hypothesis) cube B_2 is shown as the average of five 2-dim surfaces. This cube has 6 class regions. The dotted boxes show approximately the size of dual rectangular windows (see text) in proportion to target size, region stripes' sizes, and cube's spatial area $R \times C = 256 \times 256$ pixels. There are no targets in B_2.
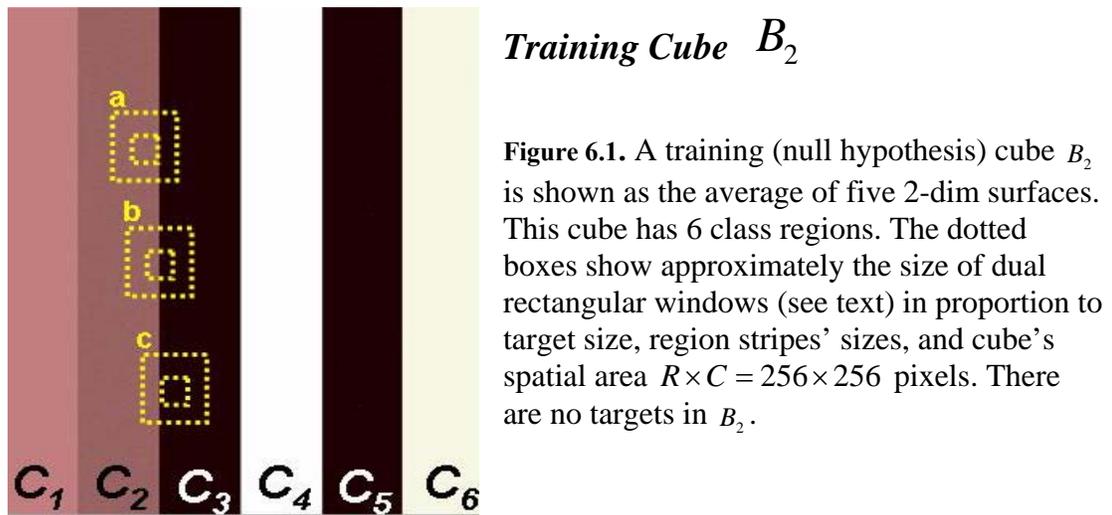
configurations will be discussed later in Subsection 6.3.2), where the repetition index g was dropped for convenience. In Fig. 6.1, the five vector components of each

sample within C_k $(k=1,\cdots,6)$ are averaged out and presented in this figure as a pixel—see also (2.5).

Anomaly detectors will test an idealized MS cube using the so-called dual rectangular window mechanism (see such a dual rectangular window at locations labeled as **a**, **b** and **c** in Fig. 6.1), where a sample $\mathbf{W}_1 \in \mathbf{R}^{K \times n_1}$ (a K dimensional column vector of sample size $n_1 = n^2$) is constructed from an $n \times n$ block of data within the idealized MS cube [see, for instance, the inside or interior window at location **a** in Fig. 6.1, and also (2.9)], and tested against another sample $\mathbf{W}_2 \in \mathbf{R}^{K \times n_2}$ (a K dimensional column vector of sample size n_2) that is constructed from data immediately surrounding this $n \times n$ block (see the outside window at location **a** in Fig. 6.1). The entire cube can be tested in this context by systematically testing all overlapping blocks of data (a pixel movement at a time to the right and/or down starting from the upper left hand corner) against their immediate surrounding samples across the spatial area of this cube. After testing the entire cube, these detectors produce two dimensional output surfaces; see, for instance, the RX multivariate detector's output surface format in (4.9) and the AVT univariate detector's output surface format in (4.44), where $\mathbf{W}_1$ (reconstructed sample from the inside window at a given location in the cube) and $\mathbf{W}_2$ (reconstructed samples from the outside window from that given location) correspond to $\mathbf{W}_1$ and $\mathbf{W}_2$ used in the detectors' formulas in Chapter 4.

Notice in Fig 6.1 that as the dual rectangular window is positioned at different locations in the cube, the outside window (and/or the inside window) may have

samples belonging to different classes. This particular case will be referred to herein as a *mixture* of different classes and denoted by $M(\bullet)$. For instance, a hypothetical multivariate sample $\left(C_2^{(g,23,1)}, \cdots, C_2^{(g,23,50)}, C_3^{(g,24,1)}, \cdots, C_3^{(g,24,50)}\right)$, or $\left(C_3^{(g,24,1)}, \cdots, C_3^{(g,24,50)}, C_2^{(g,23,1)}, \cdots, C_2^{(g,23,50)}\right)$, would be denoted by $M(C_1, C_2)$ and another hypothetical sample $\left(C_1^{(g,1,1)}, \cdots, C_1^{(g,1,40)}, C_2^{(g,2,1)}, \cdots, C_2^{(g,2,30)}, C_5^{(g,3,1)}, \cdots, C_5^{(g,3,30)}\right)$ would be denoted by $M(C_1, C_2, C_5)$.

Shifting the attention to a testing (alternative hypothesis) cube $\mathbf{BT}_z^{(g)}$, this cube is constructed by using an independently constructed cube $\mathbf{B}_h^{(g)}$ and embedding independently generated target random samples in this cube (see an example in Fig. 6.2, and construction details in Subsection 6.3.3).
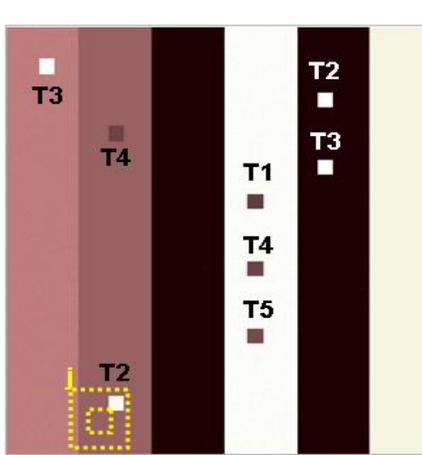
Testing Cube BT_2

Figure 6.2. A testing (alternative hypothesis) cube BT_2 is shown as the average of five 2-dim surfaces. This cube was constructed by using as baseline an independently constructed cube B_2 and embedding independently generated target samples (e.g., T3). By design, spatial areas representing squared targets are equal to the spatial area of the inside square window.

Target random samples are also i.i.d. normal, or

$$\left\{T_w^{(g,r,c)}\right\}_{r=i,c=j}^{I,J} \sim \quad i.i.d. \quad N(\tau_w, \Xi);$$ (6.3)

118

where $T_w^{(g,r,c)}$ and τ_w are column vectors $\left(T_w^{(g,r,c)} \in \mathbf{R}^K; \tau_w \in \mathbf{R}^K\right)$, $\Xi \in \mathbf{R}^{K \times K}$;

$w = 1, \cdots, 5$ indexes a different target class; and

$r = i, c = j$ $\left(1 \leq i, \cdots, I \leq R; 1 \leq j, \cdots, J \leq C\right)$ index the row and column, respectively, in

$\mathbf{BT}_z^{(g)}$. (Here also Ξ is shared by all target classes, thus, τ_w determines the wth target

class, where $\tau_1 \neq \tau_2 \neq \tau_3 \neq \tau_4 \neq \tau_5$, $\mu_k \neq \tau_w$ $\left(k = 1, \cdots, 6; w = 1, \cdots, 5\right)$ and $\Xi \neq \Sigma$ (i.e.,

all background classes are different from target classes).

An $n \times n$ block of data representing a target class in $\mathbf{BT}_z^{(g)}$ is labeled as $\mathrm{T}w$

($w = 1, \cdots, 5$); see, for instance, Fig. 6.1.

6.3. Simulation Plan and Construction of Cubes

We provide herein a simulation plan (Subsection 6.3.1) and the details to carry

out the simulation experiments using idealized MS cubes (Subsection 6.3.2 and

Subsection 6.3.3.)

6.3.1. Simulation Plan

The simulation experiment plan for this chapter uses the data models

described in Section 6.2 and parameter specifications made in Subsection 6.3.2 and

Subsection 6.3.3. The plan follows:

i.   From Section 6.2, null hypothesis (training) cubes $\left\{\mathbf{B}_h^{(g)}\right\}_{h=1}^3$ (or, for

convenience, $\left\{B_h\right\}_{h=1}^3$, which excludes the experiment repetition index

$g = 1, \cdots, 1500$), were constructed according to Subsection 6.3.2, using

independent realizations of (6.1), and used to determine cutoff thresholds for

the detectors. These cubes represent three spatial background configurations.

119

Training cubes B_1, B_2, and B_3 increase the difficult level for these detectors from easiest B_1 to hardest B_3 in terms of maintaining a relatively low cutoff threshold for a given type I error and fixed sample size (dual window size). A particular kind of background configuration are attained by determining where samples of similar or different classes are placed with respect to each other throughout the cube's spatial area, as described in Subsection 6.3.2; see Fig. 6.3.

ii. From Section 6.2, alternative hypothesis (testing) cubes $\left\{\mathbf{BT}_z^{(g)}\right\}_{z=1}^4$ (or, for convenience, $\left\{BTz\right\}_{z=1}^4$, which also excludes the experiment repetition index g) were constructed according to Subsection 6.3.3 by first generating independent realizations of $\mathbf{B}_h^{(g)}$, as described in (i), and then replacing at predetermined spatial locations in $\mathbf{B}_h^{(g)}$ some of the background samples $\left\{C_k^{(g,r,c)}\right\}_{r=i,c=j}^{I,J}$ [see (6.1)] with i.i.d. multivariate samples of target classes $\left\{T_w^{(g,r,c)}\right\}_{r=i,c=j}^{I,J}$ [see (6.3)]. (The term *predetermined* means that the locations where targets are found are stored so that we can determine later whether a detection is a correct one or a false positive.) Target class samples were generated independently from each other.

In this simulation experiment, $\mathbf{BT}_1^{(g)}$ uses the background configuration of $\mathbf{B}_1^{(g)}$, $\mathbf{BT}_2^{(g)}$ uses the background configuration of $\mathbf{B}_2^{(g)}$, and $\mathbf{BT}_3^{(g)}$ and $\mathbf{BT}_4^{(g)}$ use the background configuration of $\mathbf{B}_3^{(g)}$. The number of targets and their locations in the scene background are the same for $\mathbf{BT}_1^{(g)}$ and $\mathbf{BT}_2^{(g)}$, but

different from the ones in $\mathbf{BT}_3^{(g)}$ and $\mathbf{BT}_4^{(g)}$. Some of the target locations were intentionally selected in order to increase the difficulty level for the anomaly detectors, as they apply their corresponding cutoff thresholds to alternative hypothesis cubes featuring different background configurations, but a fixed dual-window size corresponding to the target size in the cube, given that all targets have the same spatial size.

iii.   A detector produces multiple trial results testing a cube, since it only tests a spatial block location at a trial and the detector is applied to the entire cube.

A detector trial then, in this simulation experiment, corresponds to comparing samples representing a block of data in the imagery to samples representing data in the outer ring of this block of data. As described in Section 6.2 (see also Subsection 2.4.2), since $\mathbf{W}_1$ denotes a constructed sequence $1, \cdots, n_1$ of all samples observed from a block of data in the cube and $\mathbf{W}_2$ denotes a constructed sequence $1, \cdots, n_2$ of samples observed from this block's outer ring, comparing $\mathbf{W}_1$ to $\mathbf{W}_2$ constitutes a trial. Samples $\mathbf{W}_1$ and $\mathbf{W}_2$ are shared by all the detectors used in this simulation experiment.

Making similar comparisons across the entire imagery for overlapping blocks of data produce multiple trial results from which a cutoff threshold and Type I can be estimated for a given detector, or power of the test can be assessed, depending on whether it is a training or testing activity.

Attaining cutoff thresholds and estimating Type I errors will be done during a training activity using cubes from (i), and power of the test will be estimated during testing activities using cubes from (ii). Sample sizes n_1 and

121

n_2, which depends on the sizes of the inside window and outside window, respectively, are fixed at once for this chapter to $n_1 = 81$ (from a 9×9 area) and $n_2 = 208$ (from the extension of 4 pixels beyond the 9×9 block).

Note that depending on which cube (imagery) is used and where in the imagery the block of data is located, $\mathbf{W}_1$ and/or $\mathbf{W}_2$ may have homogeneous or mixtures of background classes and/or target samples Tw, see Fig. 6.4. Note also that, in this simulation experiment, the proportions of different classes in a mixture are not predetermined, as they were in Chapter 5; they systematically represent change variations in all possible combinations in a mixture, as these detectors are systematically applied to all spatial locations in the given cube.

iv. Since the industry standard RX detector (see Subsection 4.2.1.3) assumes that $\mathbf{W}_1$ and $\mathbf{W}_2$ consist of i.i.d. multivariate normal samples of unknown means and unknown but equal covariance matrices, for this simulation we fixed at once covariance matrices of different background classes to Σ [see (6.1) and Subsection 6.3.2], and fixed at once covariance matrices of different target classes to Ξ [see (6.3) and Subsection 6.3.3], where the correlation parameters in both Σ and Ξ were all arbitrarily set to positive one.

v. Attaining Cutoff Thresholds: Applying a detector to a training cube B_h in (i) will produce *57,121* trial results $[(R-17)\cdot(C-17)=(256-17)\cdot(256-17)]$. These *57,121* trial results will be used to estimate an empirical PDF. A cutoff threshold will be attained using the standard quantile method on the estimated

PDF. The desired type I error probability (α_t) for the quantile method will vary, where $t = 1, \cdots, 4$ indexes a set of type I error probabilities. In this simulation, we arbitrarily set these error probabilities to four values: $\alpha_1 = 10^{-1}, \alpha_2 = 10^{-2}, \alpha_4 = 10^{-3}$, and $\alpha_4 = 10^{-4}$. This procedure will be repeated for each detector, such that, each detector will share the same samples across the spatial area of B_h. This procedure will also be repeated for each background configuration, i.e., B_1, B_2 and B_3.

vi. Estimating the Type I Error: Each detector will be applied to an independently generated cube $B_h^{(g)}$ according to (i), where each detector's corresponding cutoff threshold will be applied to the corresponding detector's output trial results in order to estimate the type I error $\hat{\alpha}_t^{(g)}$ of each detector for a given experiment repetition g. The type I error is estimated by counting the number $m_1^{(g)}$ of trials that satisfy the detector's output values being greater than the corresponding detector's cutoff threshold $\varepsilon_t^{(g)}$ and computing the ratio $\hat{\alpha}_t^{(g)} = m_1^{(g)} / m$, where $m = 57,121$. (Notice that m is sufficiently large for the estimation of the lowest type I error probability $\alpha_4 = 10^{-4}$.) This procedure is also repeated for each background configuration, i.e., $B_1^{(g)}$, $B_2^{(g)}$ and $B_3^{(g)}$.

vii. Estimating the Power (1.0 – Type II Error): A test cube will be generated according to (ii) and introduced to each detector, where each detector's corresponding cutoff threshold will be applied to the corresponding detector's output results in order to estimate the *power* $\hat{\beta}_t^{(g)}$ for a given experiment

repetition g. The type II error is estimated by counting the number $m_2^{(g)}$ of

trials that satisfy the detector's output values being lower than the

corresponding detector's cutoff threshold $\varepsilon_t^{(g)}$, and satisfy also that the trial

location in the test cube corresponds to the location of a target sample (recall

that the inserted locations of target samples are known, although these

detectors do not use that information). The type II error then can be estimated

by computing the ratio $m_2^{(g)}/m_3$, such that power $\hat{\beta}_t^{(g)} = 1.0 - \left(m_2^{(g)}/m_3 \right)$,

where m_3 is the total number of target samples that is present in the test

cube—this number is known for each test cube, although not used by the

detectors.

viii.   Estimating Performance Confidence Interval (CI): In order to check

variability using results from (vi) and (vii), confidence intervals for (vi)

$$\overline{\alpha}_t \pm z_{v/2}\sqrt{\frac{1}{G-1}\sum_{g=1}^{G}\left(\hat{\alpha}_t^{(g)} - \overline{\alpha}_t\right)^2} \text{ and for (vii) } \overline{\beta}_t \pm z_{v/2}\sqrt{\frac{1}{G-1}\sum_{g=1}^{G}\left(\hat{\beta}_t^{(g)} - \overline{\beta}_t\right)^2} \text{ will}$$

be estimated, where $z_{v/2} = \Phi^{-1}\left(1 - v/2\right)$ is the $1 - v/2$ quantile $\left(v = 0.05\right)$ of

the standard normal distribution; estimated quantities $\overline{\alpha}_t = G^{-1}\sum_{g=1}^{G}\hat{\alpha}_t^{(g)}$ and

$$\overline{\beta}_t = G^{-1}\sum_{g=1}^{G}\hat{\beta}_t^{(g)} \text{ are indexed by } g = 1,\cdots,G \text{ for } t = 1,\cdots,4 \text{ and } G = 1500.$$

Tabulated results will be shown later on for each detector using

$\alpha_1 = 10^{-1}, \alpha_2 = 10^{-2}, \alpha_3 = 10^{-3}, \alpha_4 = 10^{-4}.$

6.3.2. Background Cube Construction

This subsection presents in-depth details on the construction of idealized background training cubes $B_1^{(g)}$, $B_2^{(g)}$ and $B_3^{(g)}$, where the experiment repetition index g will be removed for convenience. In addition to parameter specifications, the information contained in this subsection and in follow-on subsections includes computer-programming perspective details for estimating power of detectors and other metrics discussed in this chapter.

As mentioned in Subsection 6.2, the *easiest* background cube— B_1 —has the same format of $\mathbf{X}$ in (2.4) with $R = C = 256$ and $K = 5$; B_1 has only samples of a single background class, or [using (6.1) and dropping the index g]

$$\left\{ C_1^{(r,c)} \right\}_{r=1,c=1}^{R,C} \sim \quad i.i.d. \quad N\left(\mu_1, \Sigma\right) \tag{6.4}$$

where $\mu_1 = \begin{bmatrix} 630 \\ 640 \\ 720 \\ 660 \\ 650 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 10.0000 & 14.1421 & 20.0000 & 14.1421 & 10.000 \\ 14.1421 & 20.0000 & 28.2843 & 20.0000 & 14.1421 \\ 20.0000 & 28.2843 & 40.0000 & 28.2843 & 20.0000 \\ 14.1421 & 20.0000 & 28.2843 & 20.0000 & 14.1421 \\ 10.0000 & 14.1421 & 20.0000 & 14.1421 & 10.0000 \end{bmatrix}$.

The actual component values in μ_1 are unimportant and they are not based on spectral physics; however, we arbitrarily and intentionally selected those values shown in (6.4) to yield a concave shape on the plot of component values (vertical axis) verses their corresponding component number (horizontal axis), i.e., [650 660

720 640 630] versus [1 2 3 4 5], because hyperspectral responses from natural clutter backgrounds usually show a concave shape. The variances in the diagonal of Σ were arbitrarily set to those values shown in (6.4), and the correlations imbedded in Σ were all arbitrarily set to positive one.

Background cube B_2 consists of samples from six classes, see (6.2). In order to make these six classes different, the remainder five classes have the following parameter specifications:

$$\mu_2 = \mu_1 - \begin{bmatrix} 300 \\ 300 \\ 300 \\ 300 \\ 300 \end{bmatrix}, \ \mu_3 = \mu_1 - \begin{bmatrix} 820 \\ 780 \\ 820 \\ 780 \\ 820 \end{bmatrix}, \ \mu_4 = \mu_1 + \begin{bmatrix} 1400 \\ 1400 \\ 1400 \\ 1400 \\ 1400 \end{bmatrix}; \ \mu_5 = \mu_1 - \begin{bmatrix} 800 \\ 800 \\ 800 \\ 800 \\ 800 \end{bmatrix}, \ \mu_6 = \mu_3 + \begin{bmatrix} 2000 \\ 2000 \\ 2000 \\ 2000 \\ 2000 \end{bmatrix}, \text{or}$$

$$\mu_2 = \begin{bmatrix} 330 \\ 340 \\ 420 \\ 360 \\ 350 \end{bmatrix}, \mu_3 = \begin{bmatrix} -190 \\ -140 \\ -100 \\ -120 \\ -170 \end{bmatrix}, \mu_4 = \begin{bmatrix} 2030 \\ 2040 \\ 2120 \\ 2060 \\ 2050 \end{bmatrix}, \mu_5 = \begin{bmatrix} -170 \\ -160 \\ -80 \\ -140 \\ -150 \end{bmatrix}, \mu_6 = \begin{bmatrix} 1810 \\ 1860 \\ 1900 \\ 1880 \\ 1830 \end{bmatrix}.$$

The pixel average of five co-registered images representing B_1 is shown in Fig. 6.3 (left); this figure also shows the average representations for B_2 and B_3 (center and right, respectively).
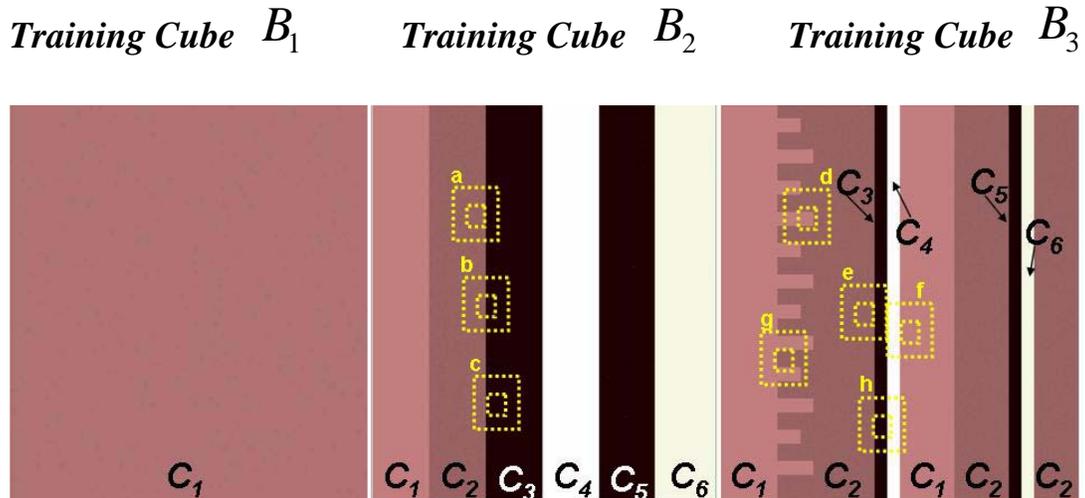
Figure 6.3. Examples of training cubes B_1, B_2, and B_3 are shown from left to right, respectively, as the average of five co-registered images per cube. The dotted boxes show approximately the size of a dual rectangular window in proportion to background stripes' sizes and cube's spatial area. Notice that depending on the window's position in a cube, the outside window may observe multivariate samples of *1* to *3* classes whereas the inside window observes only samples of a single class—see, for instance, window positions labeled as d and h.

Background cube B_3 was constructed using independent simulated realizations of the same six classes used for B_2, but B_3 displays very different spatial configurations from those in B_2. Using as reference Fig. 6.3 (right) and denoting *i1* and *i2* as the *starting* column and *ending* column, respectively, C_1 in B_3 covered columns *i1 = 1* to *i2 = 40*, having horizontal extensions of two sizes (*9* x *18*) and (*9* x *27*) located at rows (*10, 30, 50, 70, 90, 110, 130, 150, 170, 190, 210,* and *230*) for a total of *12* horizontal extensions of C_1, see B_3 in Fig. 6.1; C_2 covered columns *i1 = 41* to *i2 = 109*; C_3 covered columns *i1 = 110* to *i2 = 118*; C_4 covered columns *i1 = 119* to *i2 = 127*; C_1, being used again, covered columns *i1 = 128* to *i2 = 166*; C_2, being used again, covered columns *i1 = 129* to *i2 = 205*; C_5 covered columns *i1 =*

206 to *i2 = 214*; C_6 covered columns *i1 = 215* to *i2 = 223*; C_2, being used again,

covered columns *i1 = 224* to *i2 = 256*.

The column widths of classes $C_3, C_4, C_5,$ and C_6 in B_3 were chosen to match

the column width of the inside window—see, for instance, window position labeled

as **h** in Fig. 6.3 (right). Recall that the inside-window size was chosen to cover a *9* x *9*

spatial area, which coincides with the arbitrarily chosen target size; the size of the

outside window was chosen to cover a *17* x *17* spatial area minus the concentric area

of the inside window within the outside window. Recall also that both windows slide

concentrically across the spatial area of the given cube.

6.3.3. Background Target Cube Construction

As discussed in Section 6.2, background-target cubes (or alternative

hypothesis cubes) were constructed in order to estimate the power of correctly

detecting targets. Five target classes were arbitrarily chosen, and i.i.d. samples

representing these classes were independently generated according to (6.3), using the

following arbitrarily chosen parameter specifications:

$$\tau_1 = \mu_1 - \begin{bmatrix} 620 \\ 580 \\ 620 \\ 580 \\ 620 \end{bmatrix}, \ \tau_2 = \tau_1 + \begin{bmatrix} 2000 \\ 2000 \\ 2000 \\ 2000 \\ 2000 \end{bmatrix}, \ \tau_3 = \tau_1 + \begin{bmatrix} 2050 \\ 2050 \\ 2050 \\ 2050 \\ 2050 \end{bmatrix}, \ \tau_4 = \tau_1 + \begin{bmatrix} 50 \\ 50 \\ 50 \\ 50 \\ 50 \end{bmatrix}, \ \tau_5 = \tau_1 + \begin{bmatrix} 100 \\ 100 \\ 100 \\ 100 \\ 100 \end{bmatrix}$$

or

$$\tau_1 = \begin{bmatrix} 10 \\ 60 \\ 100 \\ 80 \\ 30 \end{bmatrix}, \tau_2 = \begin{bmatrix} 2010 \\ 2060 \\ 2100 \\ 2080 \\ 2030 \end{bmatrix}, \tau_3 = \begin{bmatrix} 2060 \\ 2110 \\ 2150 \\ 2130 \\ 2080 \end{bmatrix}, \tau_4 = \begin{bmatrix} 60 \\ 110 \\ 150 \\ 130 \\ 80 \end{bmatrix}, \tau_5 = \begin{bmatrix} 110 \\ 160 \\ 200 \\ 180 \\ 130 \end{bmatrix},$$

and

$$\Xi = \begin{bmatrix} 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \\ 100 & 100 & 100 & 100 & 100 \end{bmatrix}.$$

Targets were constructed to form sub-cubes of constant volumes $9 \times 9 \times 5$, such that, 9×9 represents the spatial area of targets, which coincides with the size of the inside window. For simplicity, the correlations values in Ξ were all equal to *1*, and the variances were all equal to *100*. Targets samples were generated independently of each other and independently of background samples.

Realizations of BT_1 were constructed by first simulating realizations of B_1 and embedding $9 \times 9 \times 5$ target sub-cubes using (6.3) in order to generate i.i.d. samples. The spatial areas of these targets are labeled in Fig. 6.4 (left) as $T1, T2, T3, T4,$ and $T5$, where Tw $(w = 1, \cdots, 5)$ corresponds to a target class using i.i.d. samples according to (6.3). As the particular locations of targets in BT_1 are unimportant, they were arbitrarily chosen to be significantly apart from each other

(e.g., *35* pixels apart from any other target's centroid, or greater) in order to avoid having two targets being observed at the same time using the given fixed-size dual window, see window blocks labeled by **i** in Fig. 6.4 (left).

Similarly, samples of cube BT_2 were formed by simulating realizations of B_2 and embedding $9 \times 9 \times 5$ target sub-cubes using (6.3), as shown in Fig. 6.4 (center). The targets were spatially collocated sufficiently apart from each other and apart from any one of the transitions of background classes (greater than *18* pixels between target centroids and a transition) in order to ensure that when the inside window happens to fully cover a target, the outside window observes samples of a single background class.   For convenience, the same targets and their locations in BT_1 were used for BT_2, see Fig. 6.4 (center).

Samples of BT_3 cubes were formed by simulating realizations of B_3 and embedding $9 \times 9 \times 5$ target sub-cubes using (6.3) for $T2, T3,$ and $T4$, as shown in Fig. 6.4 (right). The motivation here was to measure performance of anomaly detectors on a more challenging background configuration—challenging with respect to class transitions and opportunities to have samples of a single class compared to samples of two or three classes (see, for instance, window location **l** in Fig 6.4, right hand side). Targets were spatially collocated significantly apart from class transitions (greater than *25* pixels between their centroids and background transitions, as shown in Fig. 6.4.)

Samples of BT_4 cubes also use the background configuration of BT_3 (i.e., simulated realizations of B_3), but feature additional targets: $T1, T2, T3, T4,$ and $T5$,

as shown in Fig. 6.5 (right). The motivation here was to augment the testing challenge by putting some targets in transitions of different background classes, where (as the entire cube is tested by stepping by an unit the dual window across the imagery area) some of these targets will eventually be compared to two or three background classes (see, for instance, window location **a** in Fig. 6.5 [BT_4] and potential problem areas similar to window location **b** in Fig. 6.5 [BT_4]). These targets were collocated significantly apart from each other, greater than *35* pixels apart from their centroids, and, for targets that were embedded in narrow background stripes, they were put perfectly to match the width of those stripes, see Fig. 6.5 [BT_4]. Recall that the spatial size of these targets matches the spatial area of the inside window (9×9). There are some challenging areas in BT_4 cubes, for instance, the window location **b** in Fig. 6.5 (BT_4) shows a case where samples observed by the inside window belong to a single class while samples observed through the outside window belong to four classes.

In this simulation, a BT_4 cube provides the most challenging target to background configuration scenario for existing anomaly detectors, as some of background stripes' sizes correspond to the size of the inside window and, additionally, as shown in Fig. 6.5 (right), there are horizontal background extensions of vertical lengths also corresponding to the inside window's vertical length (see label **c** in BT_4).

Testing Cube BT_1 *Testing Cube* BT_2 *Testing Cube* BT_3
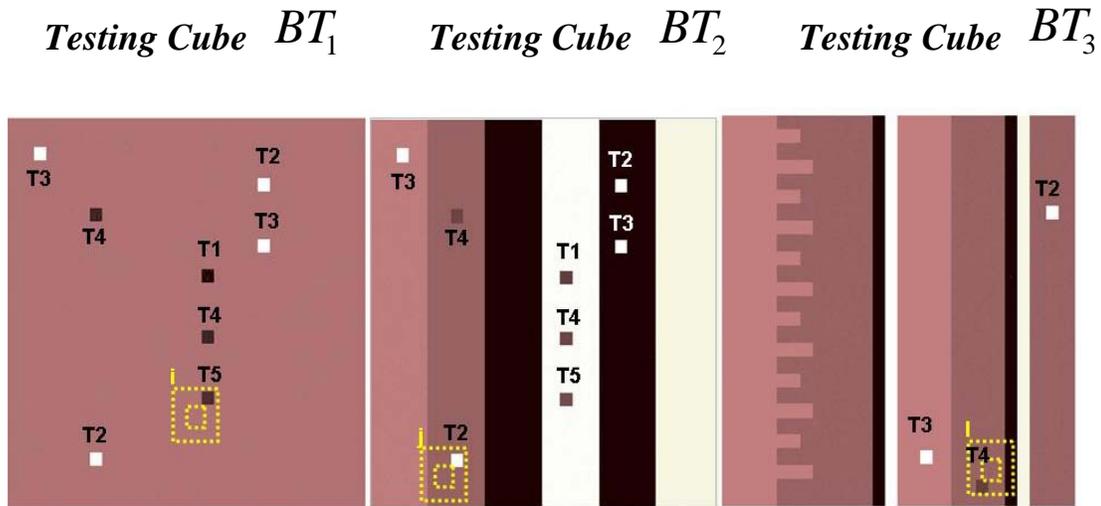
Figure 6.4. Examples of testing cubes BT_1, BT_2, and BT_3 are shown from left to right, respectively, as the average of five co-registered images per cube. The dotted boxes show approximately the size of a dual rectangular window in proportion to targets' spatial areas—see, for instance, window positions labeled as i and j. Targets are labeled according to their statistical characteristics—discussed in text.

Fig. 6.5 also depicts the *ground truth* mask—a binary image—for BT_4 (shown between B_3 and BT_4), which validates target pixel locations in the imagery (bright squares having pixel values equal to *1*). Similar *truth masks* were generated for the other testing cubes shown in Fig. 6.4. Ground truth masks are required in this context for type I. and type II error estimations, see Section 6.4.

Training Cube B_3 *Ground Truth Mask*   *Testing Cube* BT_4
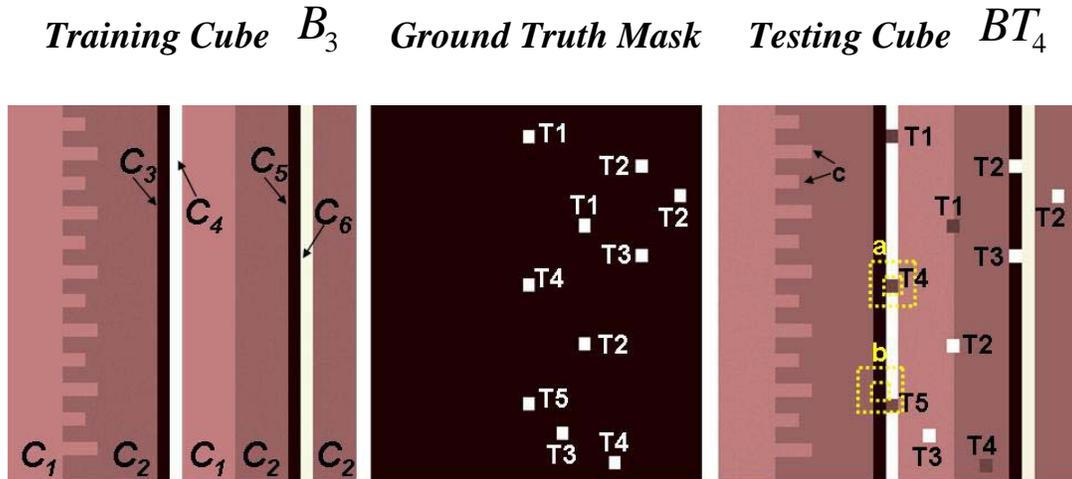
Figure 6.5. A training cube B_3, shown in far left as the average of five co-registered images, was also used to generate the background of a testing cube BT_4, see an example of BT_4 (far right) shown as the average of five co-registered images. The binary image (pixels are *one* or *zero*) shown in the center features the locations of targets in BT_4. Such a binary image is known in the target community as a ground truth mask, which is used for type I and type II error estimates. In practice, each one of the testing cube types has a corresponding ground truth mask.

6.4. Type I and Type II Error Estimations

In order to estimate type I and type II errors, two-dimensional (2D) *masks* are required to validate the spatial locations of targets in the simulated imagery. These masks are binary, i.e., the locations of target pixels are *1* and locations of background pixels are *0*; these masks are often referred to in the target community as *ground truth.*, or ground truth masks. Fig. 6.5 shows the ground truth mask for examples of BT_4. In this simulation, three ground truth masks were generated, one for BT_1 and

BT_2, as they have the same target locations; a second one for BT_3; and a third for BT_4.

In a nutshell, after a detector tests an entire simulated cube, it produces a 2D output surface of real values. If the cube is a training cube (a given background without targets), then a cutoff threshold can be chosen to yield a fixed type I error using the corresponding ground truth mask and the detector's 2D outcome surface; if the cube is a testing cube (the same background but with targets), then the detector's corresponding cutoff threshold is applied to the 2D output surface, such that, pixel values that are above the threshold and fall within target locations in the corresponding ground truth mask are labeled correct target detections; otherwise, they are labeled false detections (type II error). The power of the test is 1.0 minus the type II error. Additional details will be discussed in the next few subsections.

The type II error depends on the sample size, on the detector being used, and on the desired type I error. The sample size depends on the window size used for sampling.

Obtaining cutoff thresholds for these detectors are discussed next, followed by a discussion on measuring their type II errors.

6.4.1. Obtaining Cutoff Thresholds

A single simulated realization of the three background configurations B_1, B_2, and B_3 (see Subsection 6.3.2) were used to obtain cutoff thresholds based on the following set of chosen Type I errors α:

$$\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$$
$$= (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}).$$

<div style="text-align: right;">(6.5)</div>

(For the purpose of anomaly detection, Type I errors that are higher than 10^{-1} have no practical value).

As described in Subsection 6.3.1, the sample size of $\mathbf{W}_1$ (reconstructed samples from the 9×9 inside window) and $\mathbf{W}_2$ (reconstructed samples from the outer window) were fixed to $n_1 = 81 = 9^2$ and $n_2 = 208$, respectively. (See Section 6.2 and Chapter 4 for explanation on how to apply an anomaly detector to a data cube, using the inside-outside dual window, in order to yield a 2D output surface.)

Eight detectors were applied to the 3 simulated cubes $(B_1, B_2, \text{ and } B_3)$, using the data transformation described in Section 3.2 for the univariate detectors (AsemiP, AVT, and ANOVA), and no data transformation for the multivariate detectors (RX, FLD, EST, and DPC). Using the simulation plan discussed in Subsection 6.3.1, this procedure yielded *3* sets of cutoff thresholds per detector per chosen type I error. A corresponding set of thresholds then per detector was obtained for B_1, B_2, and B_3 based on the desired type I errors shown in (6.5). For a given type I error, one would expect these thresholds to increase as a function of increasing background complexity among B_1, B_2, and B_3. For illustration, Table 6.1 shows these sets corresponding to detectors RX (the industry standard multivariate detector) and AsemiP (univariate detector).

In reference to Table 6.1 and Table 6.2, it is worth mentioning that a detector that is less *sensitive* to these three background cubes would be more desirable for the purpose of anomaly detection. This sensitivity can be noticed by observing the rate of increase among the cutoff thresholds for a given type I error among detectors.

Table 6.1. Cutoff thresholds produced by the industry standard RX anomaly detector using its corresponding *57,121* output results per simulated background cube, $\mathbf{B}_h$ $(h = 1, \cdots, 3)$. (Note: Applying a detector across the spatial area of each cube, given the fixed dual window size, produced *57,121* output results representing a 2D output surface per cube.)

Type I Error α	Background-Only Simulated Cubes		
	B_1	B_2	B_3
10^{-1}	2.73586756020100	17.14910687789102	30.75813296377958
10^{-2}	6.69480804231700	29.87601546556457	110.93641256094044
10^{-3}	10.81913096696389	34.61656634779293	164.77893286133680
10^{-4}	16.23242322101338	41.66426583843549	284.76027446709963

Table 6.2. Cutoff thresholds produced by the AsemiP anomaly detector using its corresponding *57,121* output results per simulated background cube, $\mathbf{B}_h$ $(h = 1, \cdots, 3)$.

Type I Error α	Background-Only Simulated Cubes		
	B_1	B_2	B_3
10^{-1}	0.10714023252450	4.04119537549038	5.41993634110348
10^{-2}	0.65071118087721	15.31063919196472	19.20618525917164
10^{-3}	1.66063008313870	15.97474497597062	43.95133931745743
10^{-4}	2.60835114791086	16.31042775997816	44.96548124318737

Sets of corresponding cutoff thresholds per detector were used to estimate type I and type II errors as described next.

6.4.2. Estimating Type I and Type II Errors

This subsection presents a computer programming implementation version for estimating Type I and Type II errors in the context of this simulation experiment. As described in our simulation plan (Subsection 6.3.1), Type I errors were estimated for each detector using their corresponding sets of cutoff thresholds on their output surfaces after applying each detector to G simulated realizations of BT_1, BT_2, BT_3, and BT_4, such that, cutoff thresholds obtained using B_1 were only used to estimate type I errors on BT_1; cutoff thresholds obtained using B_2 were only used to estimate type I errors on BT_2; and cutoff thresholds obtained using B_3 were only used to estimate type I errors on BT_3 and BT_4, as these target-background cubes shared the same background configuration of B_3.

A *generic* null hypothesis H_0 can be stated for this simulation as follows: At any given location in a simulated cube, samples observed thru the inside window, $\mathbf{W}_1$, belong to the same population of samples observed thru the outside window, $\mathbf{W}_2$. This test will be repeated across the spatial area of the simulated cube, generating an output surface for each detector. The whole process is repeated using $G = 1500$ independent simulated cubes so that confidence intervals can be estimated for types I and type II errors, as discussed in Subsection 6.3.1.

The specific steps taken to obtain empirical results for type I and type II errors are shown below for a test cube, in this case BT_4, using a detector.

- Since the dual rectangular window covers an inside area of 9×9 (the area of all targets in this simulation experiment) and an outer area of

$(17 \times 17) - (9 \times 9)$, the output surface produced by a detector using simulated

cubes of fixed spatial area (256×256) will be $(256 - 17) \times (256 - 17)$

- Using a standard downsampling method, reduce the total area of the

  corresponding ground truth mask (see Fig. 6.5) from 256×256 to

  $(256 - 17) \times (256 - 17)$ in order to have the total area of this mask coinciding

  with the total area of the detector's output surface. Denote the downsampled

  ground truth mask as TRUTH.

- For $g = 1$ to $G = 1500$ (G, maximum number of repetitions)

  - Generate a simulated realization of BT_4, as described in Subsection 6.3.3

  - For $i = (1+17)$ to $(256-17)$

    - For $j = (1+17)$ to $(256-17)$

      - Using (i,j) to index the upper-left corner of the inside window [see

        (2.8), where BT_4 corresponds to $\mathbf{X}$], apply a detector (e.g., RX,

        AVT) using constructed sequences $\mathbf{W}_1$ (samples observed thru the

        inside window) and $\mathbf{W}_2$ (samples observed thru the outside

        window) as input [see, for instance, (4.8) or (4.41)]

      - This process will generate a 2D output surface OUTPUT

  - Intermediate Result: OUTPUT of size $(256 - 17) \times (256 - 17)$

  - For $t = 1$ to 4 (the maximum number of cutoff thresholds per detector)

- Let $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_4)$ be the detector's cutoff thresholds corresponding to the desired type I errors α [see (6.5)], where ε_t has a one to one correspondence with α_t. Apply ε_t to OUTPUT

- Let $v(i_0, j_0)$ be the value of a pixel located at $(i = i_0, j = j_0)$ in OUTPUT, Θ be the set of target pixel locations in TRUTH, and Θ^c the mutually exclusive set representing background pixel locations in TRUTH; notice that the set of all pixel locations in TRUTH is $\Theta \cup \Theta^c$, where $\cup$ is the union of sets. Denote N_{Θ^c} the total number of locations in Θ^c.

- A **type I error** is committed when $v(i_0, j_0) > \varepsilon_t$ and $(i_0, j_0) \in \Theta^c$, where $\in$ denotes *belongs to*. Add all instances when the type I error was committed, denoting this sum N_{fa}.

  - Estimate type I error for each α_t at a given g, or

$$\hat{\alpha}_t^{(g)} = \frac{N_{fa}}{N_{\Theta^c}} \qquad (6.6)$$

- A **type II error** is committed when $v(i_0, j_0) < \varepsilon_t$ and $(i_0, j_0) \in \Theta$. Add all instances when the type II error was committed, such that, multiple instances of the same target will be counted **only once** to avoid redundancies. (For instance, if among

139

ten targets, one of them did not have at least a portion of its spatial

area detected, the estimated type II error would be 1 divided by 10,

or 0.1; on the other hand, if a target yields a response resembling a

relatively wide peak, this target would be counted as a single

detection, as long as a portion of the peak's footprint coincides

with the target's expected spatial location. This procedure is

widely practiced in the target community because targets often

produce adjacent artifact responses. Output surfaces will be shown

later to clarify this point.)

- Denote the sum of targets that committed type II errors as N_{miss},

  and the total number of individual targets in TRUTH as N_{total} (for

  BT_4, $N_{total} = 10$, see Fig. 6.5); notice that $N_{miss} \leq N_{total}$, where

  $\leq$ is *less or equal to*

  - Estimate the *power* of the correct target detections for each

    α_t at a given g, and denote this estimate $\hat{\beta}_t^{(g)}$. The

    estimated power is 1.0 minus the estimated type II error, or

    $$\hat{\beta}_t^{(g)} = 1.0 - \frac{N_{miss}}{N_{total}} \tag{6.7}$$

  - Intermediate results at a given *g*:

    $$\hat{\alpha}^{(g)} = \left(\hat{\alpha}_1^{(g)}, \hat{\alpha}_2^{(g)}, \hat{\alpha}_3^{(g)}, \hat{\alpha}_4^{(g)} \right)$$
    $$\hat{\beta}^{(g)} = \left(\hat{\beta}_1^{(g)}, \hat{\beta}_2^{(g)}, \hat{\beta}_3^{(g)}, \hat{\beta}_4^{(g)} \right) \tag{6.8}$$

- Intermediate results using (6.8) as $g = 1, \cdots, 1500$:

$$\hat{\alpha} = \left(\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)}, \cdots, \hat{\alpha}^{(1500)} \right)$$
$$\hat{\beta} = \left(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \cdots, \hat{\beta}^{(1500)} \right) \quad (6.9)$$

- For $t = 1$ to 4 (total number of desired type I error values)

  - Estimate both type I error mean and power mean using results from all 1,500 experiment repetitions, which the tth cutoff threshold was used,

$$\hat{\mu}_t^{(0)} = \frac{1}{1500} \sum_{g=1}^{1500} \hat{\alpha}_t^{(g)}$$
$$\hat{\mu}_t^{(1)} = \frac{1}{1500} \sum_{g=1}^{1500} \hat{\beta}_t^{(g)} \quad (6.10)$$

  - Estimate type I error variance and power variance using results from all 1,500 experiment repetitions, which the tth cutoff threshold was used,

$$\hat{\sigma}_t^{(0)} = \frac{1}{1500} \sum_{g=1}^{1500} \left(\hat{\alpha}_t^{(g)} - \hat{\mu}_t^{(0)} \right)^2$$
$$\hat{\sigma}_t^{(1)} = \frac{1}{1500} \sum_{g=1}^{1500} \left(\hat{\beta}_t^{(g)} - \hat{\mu}_t^{(1)} \right)^2 \quad (6.11)$$

  - Estimate 95% confidence interval using results from all 1,500 experiment repetitions, which the tth cutoff threshold was used,

$$\left(\hat{\mu}_t^{(0)} - 1.96 \sqrt{\frac{\hat{\sigma}_t^{(0)}}{1500}}, \ \hat{\mu}_t^{(0)} + 1.96 \sqrt{\frac{\hat{\sigma}_t^{(0)}}{1500}} \right)$$

$$\quad (6.12)$$

$$\left(\hat{\mu}_t^{(1)} - 1.96 \sqrt{\frac{\hat{\sigma}_t^{(1)}}{1500}}, \ \hat{\mu}_t^{(1)} + 1.96 \sqrt{\frac{\hat{\sigma}_t^{(1)}}{1500}} \right)$$

- Final Results: 95% confidence intervals for the power of the test, i.e.,

- Type I Error Confidence Intervals

$$
\begin{aligned}
&\left(\hat{\mu}_1^{(0)} - 1.96\sqrt{\frac{\hat{\sigma}_1^{(0)}}{1500}}, \ \hat{\mu}_1^{(0)} + 1.96\sqrt{\frac{\hat{\sigma}_1^{(0)}}{1500}} \right) \\
&\left(\hat{\mu}_2^{(0)} - 1.96\sqrt{\frac{\hat{\sigma}_2^{(0)}}{1500}}, \ \hat{\mu}_2^{(0)} + 1.96\sqrt{\frac{\hat{\sigma}_2^{(0)}}{1500}} \right) \\
&\qquad\qquad\qquad \vdots \\
&\left(\hat{\mu}_4^{(0)} - 1.96\sqrt{\frac{\hat{\sigma}_4^{(0)}}{1500}}, \ \hat{\mu}_4^{(0)} + 1.96\sqrt{\frac{\hat{\sigma}_4^{(0)}}{1500}} \right)
\end{aligned}
\tag{6.13}
$$

- Power (1.0 – Type II Error) Confidence Intervals

$$
\begin{aligned}
&\left(\hat{\mu}_1^{(1)} - 1.96\sqrt{\frac{\hat{\sigma}_1^{(1)}}{1500}}, \ \hat{\mu}_1^{(1)} + 1.96\sqrt{\frac{\hat{\sigma}_1^{(1)}}{1500}} \right) \\
&\left(\hat{\mu}_2^{(1)} - 1.96\sqrt{\frac{\hat{\sigma}_2^{(1)}}{1500}}, \ \hat{\mu}_2^{(1)} + 1.96\sqrt{\frac{\hat{\sigma}_2^{(1)}}{1500}} \right) \\
&\qquad\qquad\qquad \vdots \\
&\left(\hat{\mu}_4^{(1)} - 1.96\sqrt{\frac{\hat{\sigma}_4^{(1)}}{1500}}, \ \hat{\mu}_4^{(1)} + 1.96\sqrt{\frac{\hat{\sigma}_4^{(1)}}{1500}} \right)
\end{aligned}
\tag{6.14}
$$

The results shown in (6.13) and (6.14) were computed for each detector as it tested the different kinds of simulated data cubes. Their results are shown next.

6.5. Summary of Results

This subsection presents the 95% confidence intervals that were computed for

each detector as they tested independent realizations of (alternative hypothesis) cubes

BT_1, BT_2, BT_3, and BT_4. The tables are organized such that the first column shows

the detector's name, followed by the desired type I error α_t $(t = 1, \cdots, 4)$ using (6.5),

followed by the type I error's 95% confidence intervals using (6.13), followed by the

power's 95% confidence intervals using (6.14). This organization applies to Table 6.3

through Table 6.10.

Table 6.3 shows performance results using multivariate detectors, as they tested

1,500 simulated realizations (repetitions) of BT_1 (targets in *easy* background

configuration). Performance results tabulated in Table 6.3 will be referred to herein as

calibrated performances of those multivariate detectors, since BT_1 represents the

easiest target-background configuration. The computation of individual power and

type I error estimates used 57,121 trials (window locations in the imagery) per

repetition. The same sample sequences per trial ($\mathbf{W}_1$ and $\mathbf{W}_2$) were shared by the

univariate detectors, which produced the results shown in Table 6.4. Table 6.4 shows

calibrated performance results using univariate detectors to test *1,500* simulated

realizations (repetitions) of BT_1, where estimations of power and type I error per

simulation repetition used *57,121* trial results.

It is evident from Table 6.3 and Table 6.4 that all the anomaly detectors produce

a perfect power of the correct target detection, as expected, since BT_1 cubes are

formed using a single homogeneous background class, which is very distinct from

Table 6.3. Multivariate Detection Calibrated Performance—type I error and power performances using *57,121* trials (window locations) per independent repetition of background cube $B_1^{(g)}$, where g indexes repetitions $\left\{B_1^{(g)}\right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_1^{(g)}$, $\left\{BT_1^{(g)}\right\}_{g=1}^{1500}$.

Detectors	α_t	Single Homogeneous Background Region Plus *8* Targets Easy background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
RX	10^{-1}	0.155322	0.158816	1.000000	1.000000
	10^{-2}	0.038951	0.039978	1.000000	1.000000
	10^{-3}	0.012227	0.012741	1.000000	1.000000
	10^{-4}	0.002292	0.002528	1.000000	1.000000
FLD	10^{-1}	0.170777	0.173506	1.000000	1.000000
	10^{-2}	0.087272	0.087644	1.000000	1.000000
	10^{-3}	0.078813	0.078926	1.000000	1.000000
	10^{-4}	0.077659	0.077706	1.000000	1.000000
EST	10^{-1}	0.170015	0.172612	1.000000	1.000000
	10^{-2}	0.077576	0.078367	1.000000	1.000000
	10^{-3}	0.069210	0.069504	1.000000	1.000000
	10^{-4}	0.067697	0.067767	1.000000	1.000000
DPC	10^{-1}	0.171855	0.175635	1.000000	1.000000
	10^{-2}	0.087120	0.088003	1.000000	1.000000
	10^{-3}	0.078594	0.078854	1.000000	1.000000
	10^{-4}	0.077440	0.077517	1.000000	1.000000

target classes. In many of the error results, however, the desired type I errors

$\alpha_t \left(t=1,\cdots,4\right)$ do not fall within their corresponding confidence intervals. The reason

is that since the cutoff thresholds were obtained from a background-only simulated

cube (in this case B_1), it produced lower threshold values compared to detectors'

artifact responses that can be relatively high owing to the presence of targets in a test

cube (in this case BT_1). These artifacts occur when homogenous samples in the inside

window are compared to a mixture of target-background samples in the outside

window—see, for instance, window location **i** in Fig. 6.4 (BT_1), and examples of

output surfaces' 3D plots in Fig. 6.6.

Table 6.4 Univariate Detection Calibrated Performances—type I error and power performances using *57,121* trials (window locations) per independent repetition of background cube $B_1^{(g)}$, where *g* indexes repetitions $\{B_1^{(g)}\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_1^{(g)}$, $\{BT_1^{(g)}\}_{g=1}^{1500}$.

Detectors	α_t	Single Homogeneous Background Region Plus *8* Targets Easy background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
AsemiP	10^{-1}	0.171441	0.174740	1.000000	1.000000
	10^{-2}	0.072887	0.073796	1.000000	1.000000
	10^{-3}	0.048219	0.048483	1.000000	1.000000
	10^{-4}	0.037595	0.037807	1.000000	1.000000
SemiP	10^{-1}	0.022617	0.022708	1.000000	1.000000
	10^{-2}	0.001783	0.001813	1.000000	1.000000
	10^{-3}	0.000579	0.000589	1.000000	1.000000
	10^{-4}	0.000579	0.000589	1.000000	1.000000
AVT	10^{-1}	0.156213	0.159702	1.000000	1.000000
	10^{-2}	0.055914	0.056730	1.000000	1.000000
	10^{-3}	0.039220	0.039370	1.000000	1.000000
	10^{-4}	0.039220	0.039370	1.000000	1.000000
ANOVA	10^{-1}	0.121679	0.125269	1.000000	1.000000
	10^{-2}	0.010971	0.011928	1.000000	1.000000
	10^{-3}	0.001161	0.001412	1.000000	1.000000
	10^{-4}	0.000086	0.000147	1.000000	1.000000

Fig. 6.6 shows examples of relatively high artifact responses, where 3D views

of some of the detectors' output surfaces are exhibited. (The surfaces depicted in Fig.

6.6 are examples of the intermediate results denoted as OUTPUT in Subsection 6.4.2 for different detectors.)

Table 6.5 shows tabulated results using the same multivariate detectors, but in this experiment they tested data cubes having a more challenging background configuration—a moderate target-background configuration (BT_2). The number of trials (window locations) per cube and repetitions are identical to the results presented in Table 6.3.

Tabulated results in Table 6.5 already show some signs of performance losses (higher type II error) for three of the algorithms, they are: EST, DPC, and FLD. The fundamental cause for these losses is the detectors' inability to handle transitions of regions. Recall that the cutoff thresholds used to test BT_2 were obtained using a moderate background configuration B_2, which is a six-class background-only cube. As it can be observed in Fig. 6.7, the transitions among these distinct classes yield relatively high responses using these detectors, which in turn yield relatively high cutoff thresholds for the chosen type I errors. Fig. 6.7 depicts, for instance, that some of the target responses using these detectors are comfortably above these detectors' responses on transitions of distinct background classes, and it also shows target examples that cannot respond as high. The EST detector suffered the worst performance loss testing examples of BT_2. Fig 6.8 shows two 3D viewing perspectives of the same output surface produced by the EST detector testing a single simulated realization of BT_2. Notice that the same eight targets that could be detected

146
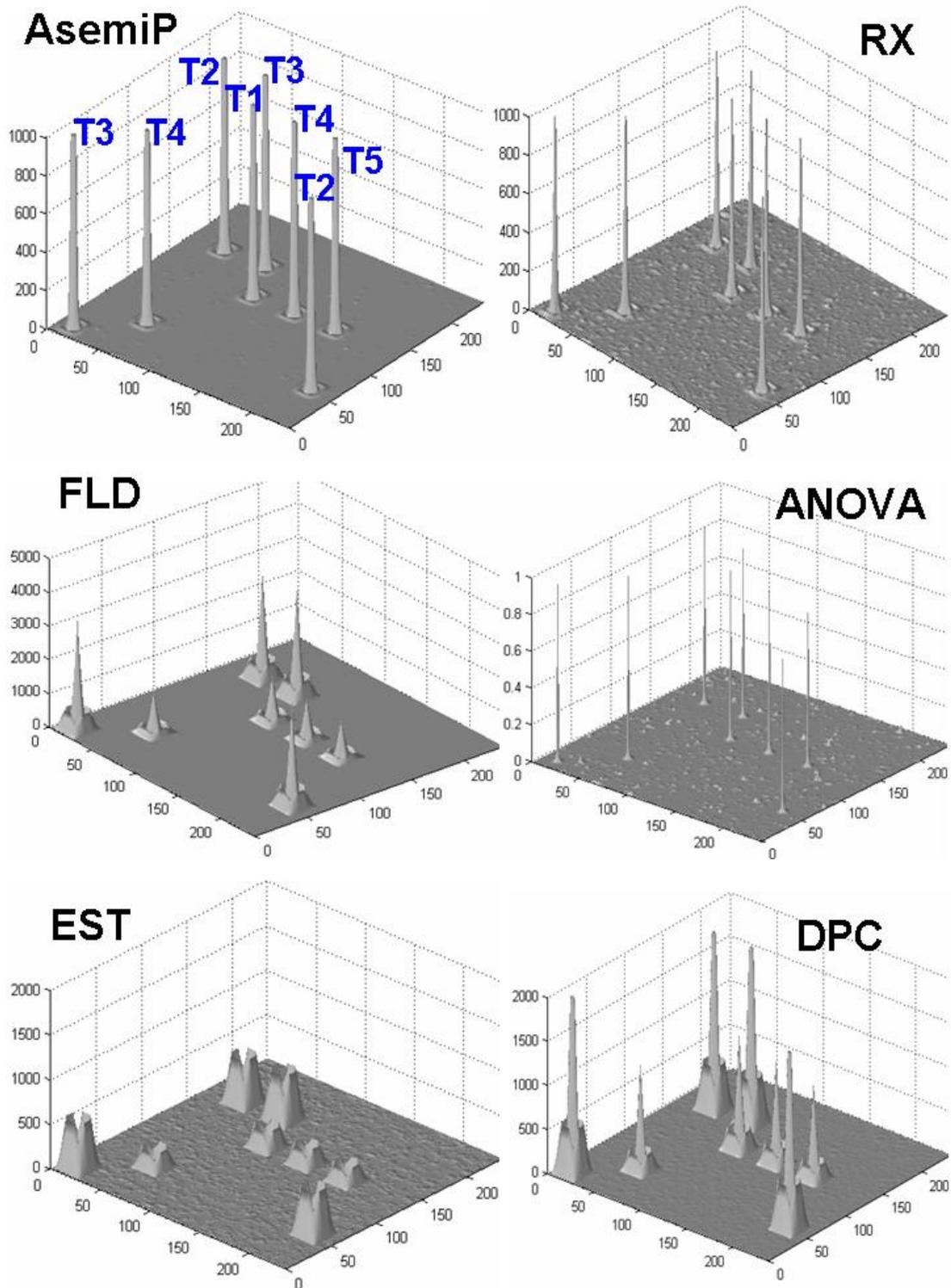
Figure 6.6. Examples of intermediate result OUTPUT, as described in text. The peaks are responses from the eight targets as *seen* by the different detectors testing a single simulated realization of BT_1. Notice the artifact responses in the vicinity of these peaks. Those artifacts contribute to the type I error, thus, increasing its estimate in respect to the desired type I error.

(calibrated performances) in BT_1 (see Fig. 6.5, Fig. 6.6 and Table 6.3), responded

below the five cutoff thresholds corresponding to the desired type I errors in (6.5) for

this detector on BT_2, thus, producing *zero* target detection—or equivalently a type II

error of *unity*, as shown in Table 6.5. The FLD and DPC detectors missed *1* out of *8*

Table 6.5. Multivariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_2^{(g)}$, where g indexes repetitions $\left\{ B_2^{(g)} \right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_2^{(g)}$, $\left\{ BT_2^{(g)} \right\}_{g=1}^{1500}$.

Detectors	α_t	Six Homogeneous Background Regions Plus *8* Targets Moderate background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
RX	10^{-1}	0.105908	0.106425	1.000000	1.000000
	10^{-2}	0.011534	0.011852	1.000000	1.000000
	10^{-3}	0.001980	0.002107	1.000000	1.000000
	10^{-4}	0.000258	0.000346	1.000000	1.000000
FLD	10^{-1}	0.107706	0.107727	0.875000	0.875000
	10^{-2}	0.005343	0.005670	0.875000	0.875000
	10^{-3}	0.001772	0.001783	0.875000	0.875000
	10^{-4}	0.001767	0.001768	0.875000	0.875000
EST	10^{-1}	0.106490	0.106511	0.000000	0.000000
	10^{-2}	0.005344	0.005670	0.000000	0.000000
	10^{-3}	0.001772	0.001783	0.000000	0.000000
	10^{-4}	0.001768	0.001768	0.000000	0.000000
DPC	10^{-1}	0.107706	0.107727	0.875000	0.875000
	10^{-2}	0.005344	0.005670	0.875000	0.875000
	10^{-3}	0.001772	0.001783	0.875000	0.875000
	10^{-4}	0.001768	0.001768	0.875000	0.875000

targets. Although the detectors FLD, EST, and DPC already show signs of

performance losses (higher type II errors) in Table 6.5—in contrast to their calibrated

performances in Table 6.3, the RX detector handled well this moderately challenging

target-background configuration. Table 6.6 shows performance results using

univariate detectors to test *1,500* simulated realizations of BT_2.

Table 6.6. Univariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_2^{(g)}$, where g indexes repetitions $\left\{ B_2^{(g)} \right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_2^{(g)}$, $\left\{ BT_2^{(g)} \right\}_{g=1}^{1500}$.

Detectors	α_t	Six Homogeneous Background Regions Plus 8 Targets Moderate background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
AsemiP	10^{-1}	0.010186	0.010240	1.000000	1.000000
	10^{-2}	0.002429	0.002429	1.000000	1.000000
	10^{-3}	0.002429	0.002429	1.000000	1.000000
	10^{-4}	0.002429	0.002429	1.000000	1.000000
SemiP	10^{-1}	0.010248	0.010305	1.000000	1.000000
	10^{-2}	0.002450	0.002473	1.000000	1.000000
	10^{-3}	0.002450	0.002473	1.000000	1.000000
	10^{-4}	0.002449	0.002470	1.000000	1.000000
AVT	10^{-1}	0.135670	0.198334	1.000000	1.000000
	10^{-2}	0.008885	0.016029	1.000000	1.000000
	10^{-3}	0.001067	0.001435	1.000000	1.000000
	10^{-4}	0.000236	0.000259	1.000000	1.000000
ANOVA	10^{-1}	0.117101	0.119892	1.000000	1.000000
	10^{-2}	0.011954	0.012952	1.000000	1.000000
	10^{-3}	0.001187	0.001442	1.000000	1.000000
	10^{-4}	0.000154	0.000230	1.000000	1.000000
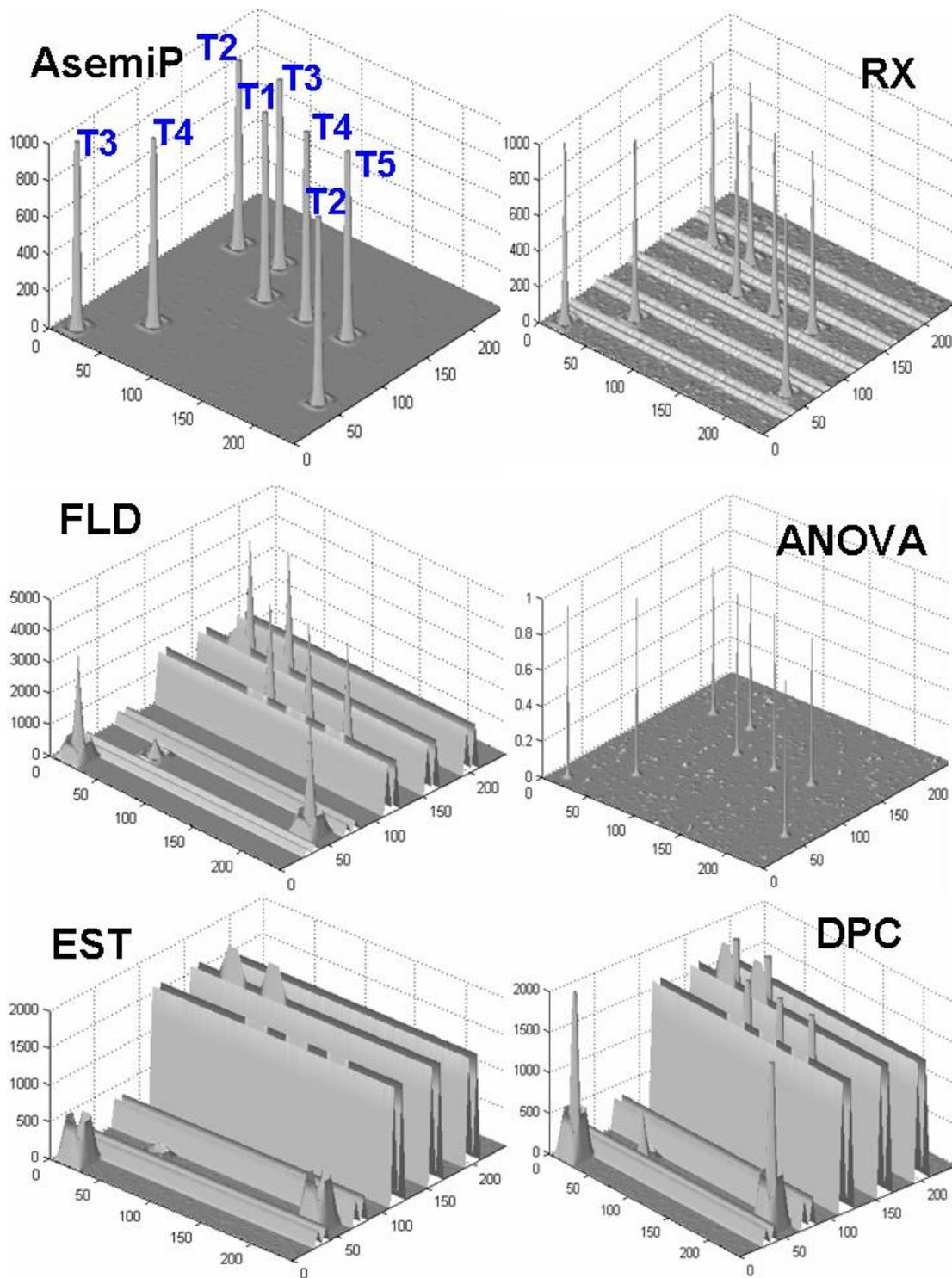
Figure 6.7. Examples of intermediate result OUTPUT, as described in text. The peaks are responses from the eight targets as *seen* by the different detectors testing a single simulated realization of BT_2. Notice, in some of these surfaces, *M* shaped row responses owing to transitions of different background classes—see, for instance, window locations a, b, and c in Fig. 6.3 (B_2). Location a yields a local peak to the left of b, b yields a local valley, and c yields a local peak to the right of b.
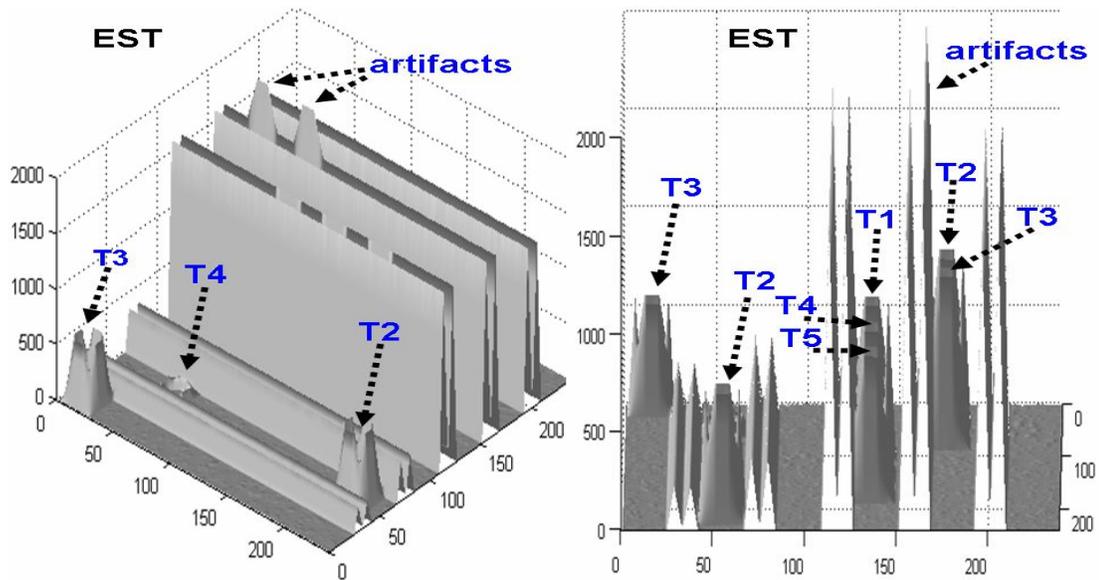
Figure 6.8. An intermediate result OUTPUT for the EST detector testing an example of BT_2. Both surfaces are the same, but shown at different viewing perspectives. Some of the artifacts shown for the view at the right hand side were due to the presence of targets T2 and T3 in the outside window—these two targets are shown immediately to the right of these artifacts. A similar case is shown for window location j in Fig. 6.4 (BT_2). The responses of all *8* targets using the EST detector were below the cutoff thresholds corresponding to this detector for this background configuration.

Some of the confidence intervals shown in Table 6.5 and Table 6.6 do not include the values of α_t, the reason for these apparent discrepancies were explained in the text discussion for Table 6.3 and Table 6.4. Univariate detectors AsemiP, SemiP, ANOVA, and AVT handled well this moderately challenging target-background configuration, see Table 6.6 and examples of output surfaces in Fig. 6.7. The output surfaces corresponding to AsemiP and ANOVA detectors in Fig. 6.7 depict how insensitivity these detectors are to those transitions of distinct classes (similar results were produced by detectors SemiP and AVT, although their surfaces are not shown in Fig. 6.7. Let's focus our attention to Table 6.7.

Table 6.7. Multivariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_3^{(g)}$, where g indexes repetitions $\{B_3^{(g)}\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_3^{(g)}$, $\{BT_3^{(g)}\}_{g=1}^{1500}$.

Detectors	α_t	Nine Homogeneous Background Regions Plus *3* Targets Difficult background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
RX	10^{-1}	0.099516	0.099962	1.000000	1.000000
	10^{-2}	0.010100	0.010353	1.000000	1.000000
	10^{-3}	0.001470	0.001488	1.000000	1.000000
	10^{-4}	0.000009	0.000032	1.000000	1.000000
FLD	10^{-1}	0.098667	0.098757	0.666667	0.666667
	10^{-2}	0.010641	0.010773	0.666667	0.666667
	10^{-3}	0.001663	0.001768	0.666667	0.666667
	10^{-4}	0.000918	0.000954	0.666667	0.666667
EST	10^{-1}	0.095054	0.095144	0.000000	0.000000
	10^{-2}	0.010278	0.010411	0.000000	0.000000
	10^{-3}	0.001663	0.001767	0.000000	0.000000
	10^{-4}	0.000918	0.000954	0.000000	0.000000
DPC	10^{-1}	0.098667	0.098757	0.666667	0.666667
	10^{-2}	0.010641	0.010773	0.666667	0.666667
	10^{-3}	0.001663	0.001767	0.666667	0.666667
	10^{-4}	0.000918	0.000954	0.666667	0.666667

Table 6.7 shows tabulated results using the same multivariate detectors, but in this case they tested data cubes having a significantly more difficult background configuration than presented by examples of BT_2, see Fig. 6.4 (BT_3 and BT_2). Table 6.8 shows performance results produced by the same univariate detectors, as they tested *1,500* simulated realizations of BT_3.

Table 6.8. Univariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_3^{(g)}$, where g indexes repetitions $\left\{B_3^{(g)}\right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_3^{(g)}$, $\left\{BT_3^{(g)}\right\}_{g=1}^{1500}$.

Detectors	α_t	Nine Homogeneous Background Regions Plus *3* Targets Difficult background configuration, targets in homogeneous areas			
		Type I Error 95% Confidence Interval		(1.0 − Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
AsemiP	10^{-1}	0.107317	0.107851	1.000000	1.000000
	10^{-2}	0.010891	0.011159	1.000000	1.000000
	10^{-3}	0.001047	0.001141	1.000000	1.000000
	10^{-4}	0.000214	0.000230	1.000000	1.000000
SemiP	10^{-1}	0.107367	0.107903	1.000000	1.000000
	10^{-2}	0.010893	0.011158	1.000000	1.000000
	10^{-3}	0.001047	0.001142	1.000000	1.000000
	10^{-4}	0.000216	0.000234	1.000000	1.000000
AVT	10^{-1}	0.100379	0.101131	1.000000	1.000000
	10^{-2}	0.009416	0.010360	1.000000	1.000000
	10^{-3}	0.001003	0.001247	1.000000	1.000000
	10^{-4}	0.000235	0.000282	1.000000	1.000000
ANOVA	10^{-1}	0.098873	0.102113	1.000000	1.000000
	10^{-2}	0.008414	0.009312	1.000000	1.000000
	10^{-3}	0.000853	0.001049	1.000000	1.000000
	10^{-4}	0.000069	0.000105	1.000000	1.000000

As discussed earlier, for a different set of targets and background configuration, Table 6.7 shows that the EST detector using its corresponding cutoff thresholds missed all *3* targets in this difficult background configuration. Similarly, using their corresponding cutoff thresholds for this difficult background configuration, Table 6.7 shows that detectors FLD and DPC missed *1* out of *3* targets. On the other hand the multivariate RX detector handled well this relatively more difficult target-background configuration.

Results in Table 6.8 also show that all the univariate detectors could also maintain a relatively low type I error and correctly detect all *3* targets.

The gradual increase of target-background configuration complexity from BT_1 to BT_3 has shown some performance losses by multivariate detectors FLD, EST, and DPC and some performance comparability between multivariate RX and all four univariate detectors (AsemiP, SemiP, AVT, and ANOVA). Table 6.9 presents the detectors' performances on cubes representing the most difficult target-background configuration in these simulation experiments—examples of BT_4.

Table 6.9 shows performance results using the same multivariate detectors, as they tested *1,500* simulated realizations of BT_4.

The results shown in Table 6.9 for *10* targets included the same *3* targets used to obtain results shown in Table 6.7. Table 6.9 shows noticeable target detection degradations compared to previous tables, with one exception—the EST detector's performance between Table 6.7 and Table 6.9. The *3* targets that were undetected by the EST detector as shown in Table 6.7 were again missed by this detector as shown in Table 6.9 (see row for EST, $\alpha_1 = 10^{-1}$). The EST detector missed those *3* out of *10* targets using its correspondent cutoff threshold for α_1, but—ironically—it could detect other targets found in more difficult locations, see locations characterized by transitions of regions in Fig. 6.5 (window locations **a** and **b** in BT_4). This irony is what motivates some users to utilize the EST detector in real HS data.

Table 6.9. Multivariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_3^{(g)}$, where g indexes repetitions $\left\{B_3^{(g)}\right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_4^{(g)}$, $\left\{BT_4^{(g)}\right\}_{g=1}^{1500}$. .

| Detectors | α_t | Nine Homogeneous Background Regions Plus *10* Targets Difficult background configuration, *7* targets in transition areas | | | |
| | | Type I Error 95% Confidence Interval | | (1.0 – Type II Error) 95% Confidence Interval | |
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
RX	10^{-1}	0.101381	0.101805	1.000000	1.000000
	10^{-2}	0.009608	0.009831	1.000000	1.000000
	10^{-3}	0.000851	0.000861	0.700000	0.700000
	10^{-4}	0.000921	0.000923	0.500000	0.500000
FLD	10^{-1}	0.101444	0.101535	0.666667	0.666667
	10^{-2}	0.010374	0.010522	0.500000	0.500000
	10^{-3}	0.001120	0.001279	0.500000	0.500000
	10^{-4}	0.000072	0.000112	0.500000	0.500000
EST	10^{-1}	0.101303	0.101394	0.700000	0.700000
	10^{-2}	0.010374	0.010522	0.300000	0.300000
	10^{-3}	0.001120	0.001279	0.300000	0.300000
	10^{-4}	0.000072	0.000112	0.300000	0.300000
DPC	10^{-1}	0.101444	0.101535	0.666667	0.666667
	10^{-2}	0.010374	0.010522	0.500000	0.500000
	10^{-3}	0.001120	0.001279	0.500000	0.500000
	10^{-4}	0.000072	0.000112	0.500000	0.500000

As mentioned earlier, anomaly detectors are known for producing relatively high responses adjacent to target locations because, at those adjacent locations, samples of homogeneous backgrounds in the inside window are compared to a mixture of samples in the outside window—this mixture may consist of samples of targets, samples of the same background class observed thru the inside window, and

from other background classes (see, for instance, window position **b** in Fig. 6.5

[BT_4].)

Table 6.10 shows corresponding results for univariate detectors testing *1,500* simulated realizations of BT_4.

Table 6.10. Univariate detection performances—type I error and power performances using *57,121* trial results per simulated background cube $B_3^{(g)}$, where *g* indexes repetitions $\left\{B_3^{(g)}\right\}_{g=1}^{1500}$, and *57,121* trial results per target-background cube $BT_4^{(g)}$, $\left\{BT_4^{(g)}\right\}_{g=1}^{1500}$.

Detectors	α_t	Nine Homogeneous Background Regions Plus *10* Targets Difficult background configuration, *7* targets in transition areas			
		Type I Error 95% Confidence Interval		(1.0 – Type II Error) 95% Confidence Interval	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
AsemiP	10^{-1}	0.111715	0.112103	1.000000	1.000000
	10^{-2}	0.011173	0.011399	1.000000	1.000000
	10^{-3}	0.001400	0.001496	1.000000	1.000000
	10^{-4}	0.000802	0.000817	1.000000	1.000000
SemiP	10^{-1}	0.111758	0.112148	1.000000	1.000000
	10^{-2}	0.011171	0.011395	1.000000	1.000000
	10^{-3}	0.001402	0.001499	1.000000	1.000000
	10^{-4}	0.000805	0.000823	1.000000	1.000000
AVT	10^{-1}	0.103537	0.104030	1.000000	1.000000
	10^{-2}	0.026165	0.026788	1.000000	1.000000
	10^{-3}	0.019580	0.019728	1.000000	1.000000
	10^{-4}	0.018893	0.018938	1.000000	1.000000
ANOVA	10^{-1}	0.100254	0.103467	1.000000	1.000000
	10^{-2}	0.009011	0.009827	0.500000	0.500000
	10^{-3}	0.000978	0.001151	0.500000	0.500000
	10^{-4}	0.000077	0.000107	0.500000	0.500000

Despite the difficult target-background configuration presented in examples of BT_4, Results in Table 6.10 shows that the univariate detectors can maintain a relatively low type I error while detecting all *10* targets in this simulation, except for the ANOVA adaptation to an univariate detector. To better appreciate the difference in performance among different detectors, see examples of output surfaces shown in Fig. 6.9.
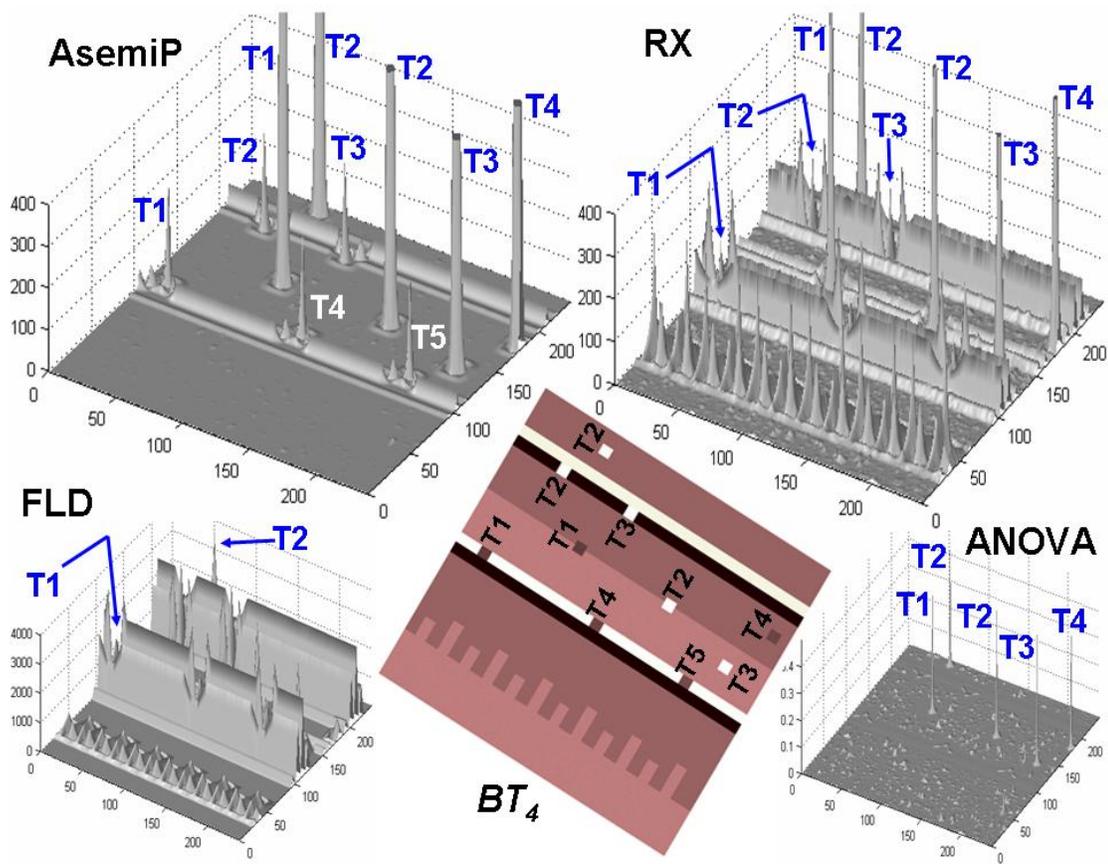
Figure 6.9. Examples of intermediate result OUTPUT, as described in text. The peaks are responses from the *10* targets as *seen* by the different detectors testing a single simulated realization of BT_4. Notice the artifact responses in the vicinity of some of these peaks. Depending on the detector, some of these artifact responses are more accentuated then targets' responses—see, for instance, the responses of targets T1, T4, T5, T2, and T3 (embedded in narrow background stripes) in the RX output surface. Those artifacts also contribute to the type I error, thus, increasing its estimate with respect to the desired type I error.

157

Fig. 6.9 shows that the univariate ANOVA detector suppresses very well the transitions of distinct background classes presented in BT_4, but unfortunately it also suppresses the most challenging locations in BT_4—where targets are present. All the five targets that were located in narrow background stripes were undetected by the ANOVA detector (see targets T1, T4, T5, T2, and T3 that are located at narrow background stripes in Fig. 6.9 [BT_4] and notice that their responses are virtually noise in the ANOVA output surface in Fig. 6.9; targets outside those narrow stripes were accentuated in this surface, see the other T1, 2 T2s, T3, and T4.) Incidentally, for displaying purposes only, the outputs surfaces shown in Fig. 6.9 for univariate AsemiP and multivariate RX were clipped at a maximum value of *400*, some of those peaks continue to significantly higher values.

Similar trend can be observed for the multivariate RX detector, see Table 6.9 and Fig. 6.9. As the desired type I error for the RX detector decreases, its estimated type II error increases. For instance, the RX responses for targets T1, T4, and T5 located at the narrow light-colored background stripe (see Fig. 6.9 [BT_4]) were undetected at the cutoff thresholds for desired type I errors $\leq \alpha_3 = 10^{-3}$, and the RX responses for targets T2 and T3 located at a second narrow background stripe (see Fig. 6.9 [BT_4] dark narrow band having T2 and T3) were also undetected at the cutoff threshold for $\alpha_4 = 10^{-4}$; thus, at $\alpha_4 = 10^{-4}$, the RX detector missed *5* out of *10* targets, see Table 6.9 (RX, row $\alpha_4 = 10^{-4}$).

Performance losses were observed earlier for the other multivariate detectors FLD, EST, and DPC, as shown in Table 6.5 and Table 6.7, now in Table 6.9.

Table 6.10 shows that, in general, the univariate detection approach handled well the most difficult target-background configuration in this simulation. The AsemiP output surface shown in Fig. 6.9 illustrates the strength of using this approach for top-view anomaly detection applications, i.e., with respect to the ten target responses seen in that output surface, the AsemiP showed that, as it tested simulated data cubes featuring a challenging target-background configuration— BT_4, it can (i) suppress window-size background region extensions of a major background class over another class, (ii) suppress challenging narrow background regions (see, for instance, window location **b** in Fig. 6.5), and (iii) suppress the local transitions of distinct background regions.

It is worth mentioning from tabulated results in Fig. 6.9 that the AsemiP detector's signal to noise ratios between target responses (signal) and background responses (noise) are not necessarily the same for all targets, or for that matter for the same target type located in different local background configurations (see, for instance, target responses of T1, T2, and T4 in Fig. 6.9, as they are spatially located at two different locations in BT_4—inside a narrow background stripe and outside this stripe. The same observation can be made for all the other detectors' performances as well. This issue relates to the ability of a detector to yield a high signal to noise ratio involving sample mixtures of different classes, which was addressed in Chapter 5. The message here is that the univariate detection approach proposed in this dissertation for top-view anomaly detection is not completely insensitive to the correspondence between local background configuration and dual window size, but it demonstrates a significant amount of insensitivity to difficult background scenarios,

which is in contrast to the performances of the multivariate approach normally used

for this application. More importantly, this contrast in performance between both

approaches has been consistent using these detectors to test real hyperspectral

imagery, as shown in Chapter 7. The results presented in Table 6.9 and Table 6.10

can be readily plotted in terms of Receiver's Operational Characteristics (ROC)

curves [35] for each detector. The corresponding ROC curves are shown in Fig. 6.10.
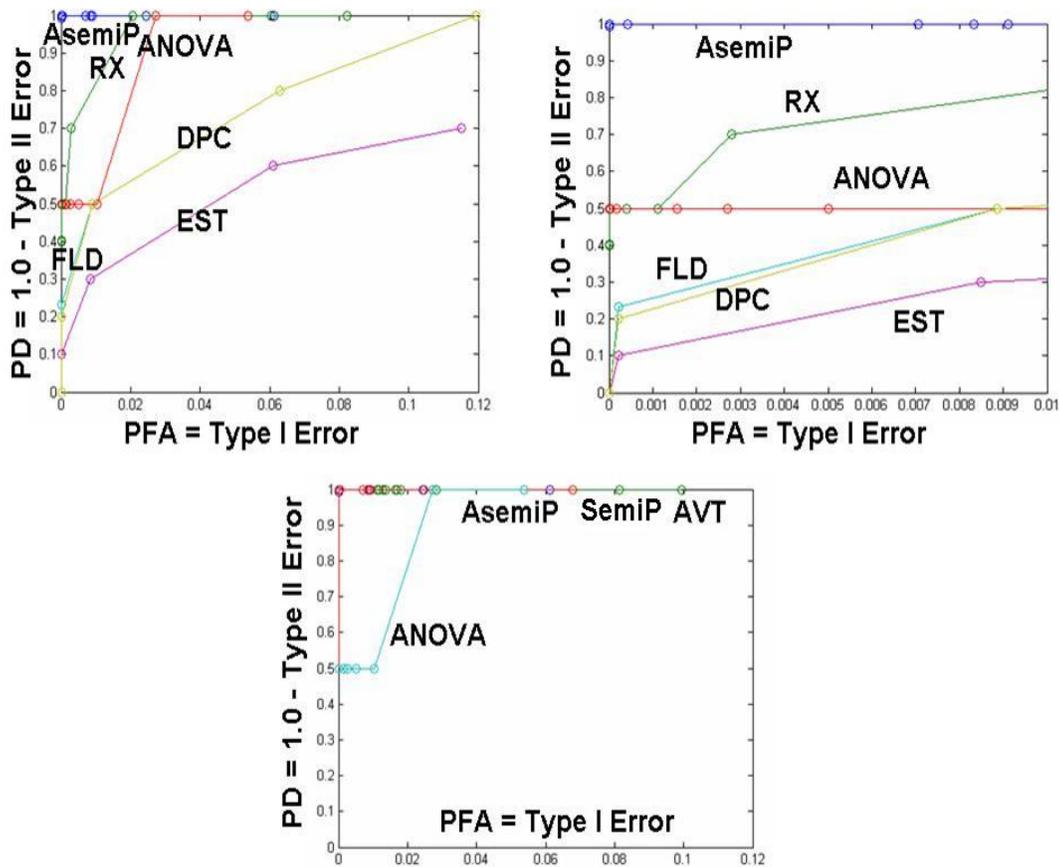
Figure 6.10. ROC curves using estimates shown in Table 6.9 and Table 6.10. The upper left side curves exhibit performances for a range between *0.0* and *0.12* (PFA) and between *0.0* and *1.0* (PD). The upper right side curves depict the same curves shown in the left but limited to a PFA range between *0.0* and *0.01*. The bottom curves include performances of detectors SemiP and AVT. (An ideal ROC curve resembles a step function starting at point [PFA = *0.0*, PD = *1.0*]).

A ROC curve plots the probability of detection (PD), which is defined as 1.0 minus the estimated type II error, versus the probability of false alarms (PFA), which is defined as the estimated type I error. The estimated mean averages of the type I and type II errors from *1,500* experiment repetitions were used to estimate PD and PFA for each detector, as they tested examples of BT_4.

The curves shown in the upper left side of Fig. 6.10 exhibit performances of univarite (AsemiP and ANOVA) detectors and multivariate (RX, FLD, EST, and DPC) detectors for a PFA range between 0.0 and 0.12 and a PD range between 0.0 to 1.0. The curves shown in the upper right side of Fig. 6.10 depicts the same curves but limited to a PFA range between 0.0 and 0.01. And the curves showed in the lower part of Fig. 6.10 show performances of all four univariate detectors AsemiP, SemiP, AVT, and ANOVA for a PFA range between 0.0 and 0.12. An ideal ROC curve resembles a step function starting at point (PFA, PD) = (0.0, 1.0).

6.6. Summary and Conclusions

This chapter has examined the performance of four two-step univariate anomaly detectors and four multivariate anomaly detectors on simulated multivariate data cubes, mimicking a target detection application from top-view anomaly detection. Results tabulated in Table 6.3 thru Table 6.9 suggest that the overall performance of univariate anomaly detectors can be significantly less dependent on, or sensitive to, the background configuration of data cubes than the overall performance of popular multivariate anomaly detectors. These univariate detectors were also able to accentuate better the presence of targets in difficult background configurations (see Fig. 6.9, Table 6.9 and Table 6.10) than multivariate detectors.

It is also worth mentioning that since the results presented in this chapter were obtained from conducting controlled simulations, estimated PD differences for these detectors could also be controlled by designed. For instance, if we increased the number of targets in regions in BT_4 having benign background configurations, the PD results of multivariate detectors—especially the RX detector—could be made to correspond to PD results produced by the univariate detectors in Table 6.10. Conversely, if we increased the number of targets in difficult local background configurations, the multivariate detectors (including the RX detector) could be made to produce significantly lower PD results than their performances shown in Table 6.9. Hence, results presented in this chapter are intended only for illustration purposes of what types of backgrounds (or background configurations) the two-step univariate detectors can perform better than popular multivariate detectors. In this illustration, the background complexity gradually increased from a relatively simple homogeneous class configuration, to a moderately difficult five region class configuration, to a difficult nine region class configuration having some of the local transitions of regions corresponding to the dual window size. Targets were strategically introduced to realizations of these background cubes, so that we could measure the effects of these background configuration changes on the anomaly detectors, as they tested these data cubes using their corresponding sets of calibrated cutoff thresholds. Performance robustness under increasing background configuration complexity, which was shown in this chapter by the univariate detectors, is highly desired in the HS research community, since real life scenarios present all kinds of unpredictable background configurations. Chapter 7 addresses ground view anomaly detection using real HS imagery.

Chapter 7   GV Anomaly Detection Using Real HS Data

7.1. Introduction

This chapter addresses the problem of anomaly detection from a ground-to-ground viewing perspective. (The notations presented in Section 2.3 and Section 2.4 apply to this chapter, unless mentioned otherwise.)

If an anomalous object (target) is defined as one made of a material that is spectrally different from all the materials composing its natural clutter background, then the question we attempt to answer in this chapter is the following: Can an algorithm suite be developed to automatically detect (or accentuate) the presence of targets in a cluttered environment, given that the imagery was recorded from the ground-to-ground viewing perspective and no prior information is known about the various materials composing the cluttered environment, the number of targets present in the scene (or if targets are present at all), the scales of targets (their relative sizes in the imagery), shapes and material types of these targets, the illumination environment, and atmospheric conditions.

Anomaly detection using GV imagery is significantly harder to address than using TV imagery, because the distances between the sensor and objects in the scene are unavailable for ground view imagery, thus, adding one more unknown variable (target scales) to the anomaly detection problem. Note, for instance, that small targets at closer range will look large—and vice versa, and multiple targets in the same scene may have different scales. The sampling method using a fixed dual rectangular window (see Chapter 6) would not be effective in this application because, in the

event that target samples are observed in the inside window, it cannot be guaranteed that the samples observed through the outside window will always belong to the clutter background. Therefore, in the event that target samples are simultaneously observed through both inside and outside windows, the detector will likely suppress the pixels representing that window location in the resulting output surface.

To circumvent target scale uncertainties, we propose to automatically take N blocks of data from random locations in the imagery and, since the targets are expected to cover a significantly small area in the imagery, label these data sets as *spectral references* of clutter background. There is, however, a probability that, if targets are present in the scene, some of these spectral reference sets will be *contaminated*, i.e., one of these spectral reference sets includes target pixels. In order to decrease the probability of contamination, we propose to repeat independently this random sampling process M number of times and will show that the probability of taking target samples by chance during these repetitions can be modeled— approximately—by the Binomial distribution family. We will use this approximation to assist on tradeoff decisions.

The remainder of this chapter is organized as follows: Section 7.2 discusses the SOC-700 HS data used for this experiment. Section 7.3 proposes a repeated (parallel) random sampling approach and models this approach by a binomial distribution; it also discusses how this sampling approach can be implemented in the context of anomaly detection and presents results using real GV HS imagery. Section 7.4 gives some insights on the detection performances shown in Section 7.3 by applying additional detectors on real GV HS imagery, but using, instead, prior information on

164

the clutter background and manual sampling to form a spectral library. Section 7.5 concludes this chapter.

7.2. Description of the SOC-700 Hyperspectral Data

The GV imagery used for this work was recorded using the SOC-700 VNIR HS spectral imager from Surface Optics Corporation, see Section 2.2. The system produces HS data cubes of dimensions $R = 640$ by $C = 640$ pixels by $K = 120$ spectral bands between 0.38 and 0.97 μm. The sensor is commercially available off the shelf [20].

Fig. 7.1 depicts samples of GV imagery recorded with the SOC-700 HS imager; each pixel in any of the four cube examples corresponds to the average of all the band (120) values at that pixel location, see (2.5).

Data cubes Cube 1, Cube 2, and Cube 3 were collected during the month of June 2004 in Fort Hunter-Liggett, California; data cube Cube 4 was collected during the month of April 2008 in Picatinny Arsenal, New Jersey. From actual ground truth, it is known that the scene in Cube 1 (see Fig. 7.1) contains three motor vehicles and a standing person in the center of that scene (i.e., two pick-up trucks to the left in proximity to each other, a man slightly forward from the vehicles in the center, and a sport utility to the right). The cluttered environment in Cube 1, Cube 2, and Cube 3 is dominated by Californian valley-type trees and/or terrain, where in Cube 2 the same sport utility vehicle and the same person stand in proximity to each other; they are located in the same valley, but at a different area from the one in Cube 2. Cube 3, although recorded in the same general geo-location of Cube 1 and Cube 2, depicts a

Cube 1 **Cube 2**

Cube 3 **Cube 4**

Figure 7.1. Examples of GV imagery. An effective GV anomaly detection algorithm suite would allow a machine to accentuate the presence of targets, while suppressing the cluttered environment, using no prior information about what constitutes clutter background or target in the imagery.

significantly more complex scenario, where, from actual ground truth, it is known that a sport utility vehicle is in the shades of a large cluster of trees. Portions of the shadowed vehicle can be observed near the center in Cube 3. Cube 4 was recorded in a wooded region in Picatinny Arsenal, New Jersey, where (according to the available

ground truth) a sport car is located behind several tree trunks and, hence, can be only partially observed in this *heavy* cluttered environment; see Fig 7.1 (left-center in Cube 4).

The four data cubes in Fig. 7.1 are independently displayed as intensity images after linear mapping the gray scale of each to the range 0-255. Pixel intensities shown in each individual surface is only relative to corresponding values in that surface; in other words, pixel values representing the same material (general terrain) may be displayed with different intensities in another surface. This fact explains, for instance, the difference in brightness between the terrains displayed in Cube 2 and Cube 3, given that both the cluttered environment and atmospheric conditions were about the same during collection of both data cubes. The strong reflections from certain parts of the vehicles captured by the sensor in Cube 1 and Cube 2 are not as dominant in Cube 3 because the vehicle in Cube 3 is in tree shades; hence, the terrain in Cube 3 is the strongest reflector in the scene.

7.3. Autonomous Sampling of the Cluttered Environment

A parallel random sampling approach is presented in this section for autonomous clutter background characterization. This approach is then incorporated into an algorithm suite in order to perform GV anomaly detection, using a favorite detector.

Results from testing this anomaly detection algorithm suite on real GV imagery are also presented in this section.

7.3.1. A Binomial Based Parallel Random Sampling Model

Assume that target pixels are present in the $R \times C$ spatial area of a $R \times C \times K$ HS data cube $\mathbf{X}$, see (2.4) and examples in Fig. 7.1, denote a the total number of target pixels in $\mathbf{X}$ and q the probability of a pixel in $\mathbf{X}$ being a target pixel out of all $A = RC$ pixels in $\mathbf{X}$, i.e., $q = \dfrac{a}{A}$. (In most applications q is unknown, and if multiple targets are present in the imagery, a will be the total number of all target pixels included in the imagery; also, these targets may or may not have the same material type.) In order to represent the unknown clutter background in the imagery, let N blocks of data—all having a fixed small area $(n \times n) << (R \times C)$—be randomly selected from the $R \times C$ area (see Fig. 7.2). In theory, for $(n \times n) = (1 \times 1)$ and using the assumption that target pixels in $\mathbf{X}$ are disjoint and randomly located across the $R \times C$ imagery area (in practice, this assumption is not satisfied when targets are present in the scene), the probability P that at least one block of data has a target pixel is

$$
\begin{aligned}
P(m \geq 1) &= p(m = 1) + p(m = 2) + \cdots + p(m = N) \\
&= 1 - p(m = 0),
\end{aligned}
\tag{7.1}
$$

where p is the binomial density function [23], given parameters q and N, and $m \in \{0, 1, \cdots, N\}$ is the number of blocks of data containing a target pixel, or

$$
p(m|q, N) = \frac{N!}{m!(N-m)!} q^m (1-q)^{N-m}.
\tag{7.2}
$$

(Symbols $|$ and ! denote *given parameters* and the *factorial operator*, respectively.)

For convenience, we will refer to $P(m \geq 1)$ as the *probability of contamination* and, consequently, *m* as the number of *contaminated* blocks of data.
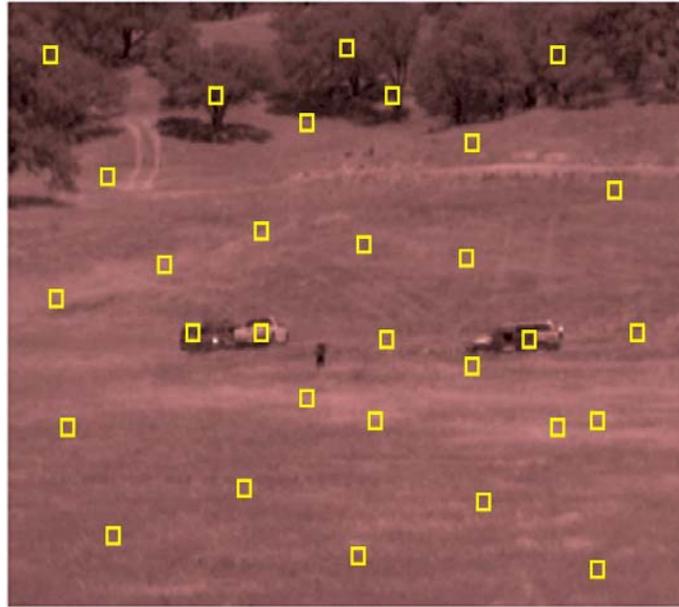
Figure 7.2. *N* small $(n \times n)$ blocks of data are randomly (autonomously) selected from the imagery $(R \times C)$ area, as spectral reference sets. In autonomous remote sensing applications, since it is unknown a priori whether target pixels are present in the imagery, a probability $P(m \geq 1)$ exists of at least a block of data being contaminated with target pixels.

The implementation of this contamination model to the autonomous background sampling problem requires that each one of the N $(n \times n)$ blocks of data be regarded as an independent reference set $\mathbf{W}_2^{(f)}$ $(f = 1, 2, \cdots, N)$ representing clutter spectra, where $\mathbf{W}_2^{(f)} \in \mathbf{R}^{K \times n_2}$ is a rearranged sequence version of the fth block

169

of data having $n_2 = n^2$ spectra. By necessity, n_2 must be significantly greater than

one—for statistical purposes—but yet significantly smaller than $A = R\,C$ (e.g.,

$\dfrac{n_2}{A} = \dfrac{20^2}{640^2} = 0.000977$) in order to be reasonable to regard a $n \times n$ block of data an

unit area on the $R \times C$ imagery area. A contaminated block of data, then, will be

treated qualitatively as a block having target pixels covering a large portion of the

block's area (e.g., greater than 0.70). In addition—when targets are present, since

pixels representing a single target are expected to be clustered in the imagery, the

assumption that each target pixel is randomly located across the imagery area will be

ignored. Using (7.1), while ignoring the non-clustered target pixel assumption,

implies that the probability of contamination will be overestimated, as blocks of data

are less likely to be randomly selected from the same cluster of target pixels. (For the

autonomous background sampling problem, it is more conservative to overestimate

the probability of contamination than to underestimate.)

Fig. 7.3 shows a plot of the probability of contamination $P(m \geq 1)$ versus N,

for two values of q (0.1 and 0.2). It is highlighted in Fig. 7.3 that, for instance, if

parameters are set to $(q, N) = (0.10, 22)$ then $P(m \geq 1) = 0.90$. Notice that for

$N = 22$, if target pixels are present but cover less than $q = 0.10$ of the imagery area,

$P(m \geq 1) = 0.90$ is overestimated by two fronts: (i) pixels from a single target are not

randomly spread across the imagery area, but clustered, and (ii) the cumulative

number of target pixels covers less than 0.10 of the imagery area. So, (7.1) provides

an upper bound (conservative) approximation of the probability of contamination,

given parameters q and N.

Fig. 7.3 also shows the trade off between having a larger number of spectral sets (increasing N) in order to adequately represent the clutter background, which is desired, and the cost of increasing probability of contamination, which is not desired. (More directly, contamination implies that once target pixels are randomly selected by chance from the imagery area, they will be used by a detector as reference set to test the entire imagery, which under the case targets would be suppressed.)
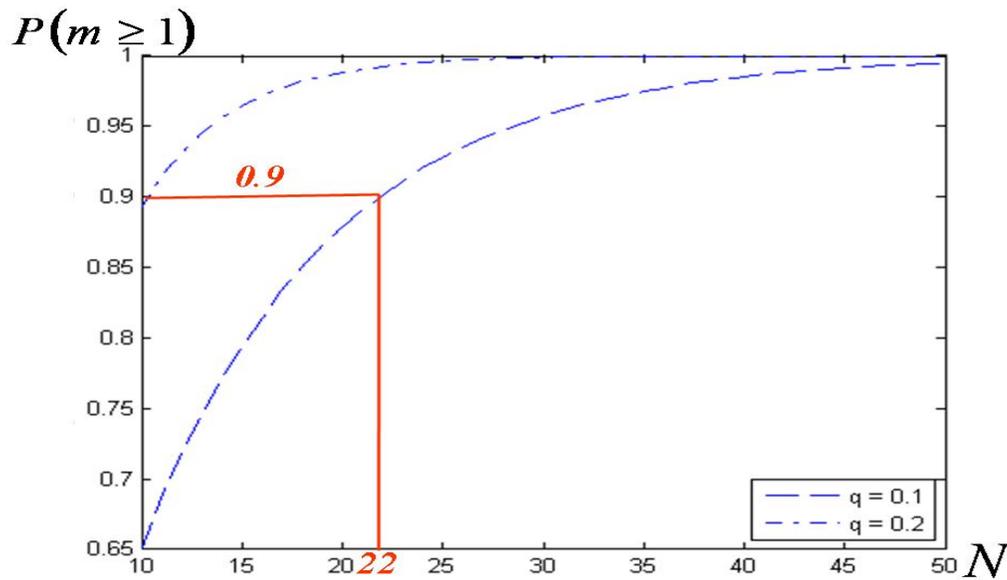
Figure 7.3. The probability $P(m \geq 1)$ of having at least a (1×1) block of data contaminated with a target pixel, as a function of N (the number of randomly selected 1×1 blocks of data), for two given values of q (the probability of randomly selecting a target pixel in the imagery area). These curves are conservative upper bounds, because target pixels are assumed to be randomly distributed across the imagery area, but in practice pixels are clustered per each target.

Since the presence of target pixels in the imagery is unknown a priori, finding a way to decrease the probability of contamination becomes a necessity. In order to decrease this probability, using an adequately large N and a sensible value for q, we propose to independently repeat the random sampling process described in this subsection M number of times. Fig. 7.4 illustrates the outcome of M repetitions. If we denote the probability of contamination of the gth random sampling process (or repetition) as $P_g(m \geq 1)$, $1 \leq g \leq M$, for a fixed q and N, note that each $P_g(m \geq 1) = P(m \geq 1)$ and, since $0.0 \leq P(m \geq 1) \leq 1.0$ and these processes will be repeated independently from each other, the overall probability $\tilde{P}$ that <u>all</u> the processes will be contaminated with at least a contaminated block of data will decrease as a function of increasing M, or

$$\tilde{P} = P_1(m \geq 1)P_2(m \geq 1)\cdots P_M(m \geq 1) = [P(m \geq 1)]^M . \tag{7.3}$$

The overall probability of contamination in (7.3) can also be expressed using the binomial distribution by letting $\tilde{m}$ be the number of independent processes that are contaminated out of M repetitions, where $\tilde{m} \in \{0,1,\cdots,M\}$, and using $P(m \geq 1)$ as the probability of contamination per process. It follows:

$$\tilde{P}(\tilde{m} = M) = \frac{M!}{\tilde{m}!(M - \tilde{m})!}[P(m \geq 1)]^{\tilde{m}}[1 - P(m \geq 1)]^{M - \tilde{m}}$$

$$= [P(m \geq 1)]^M \tag{7.4}$$

$$= \left[1 - (1 - q)^N\right]^M ,$$

for $N \cdot M \cdot q \ll A$.
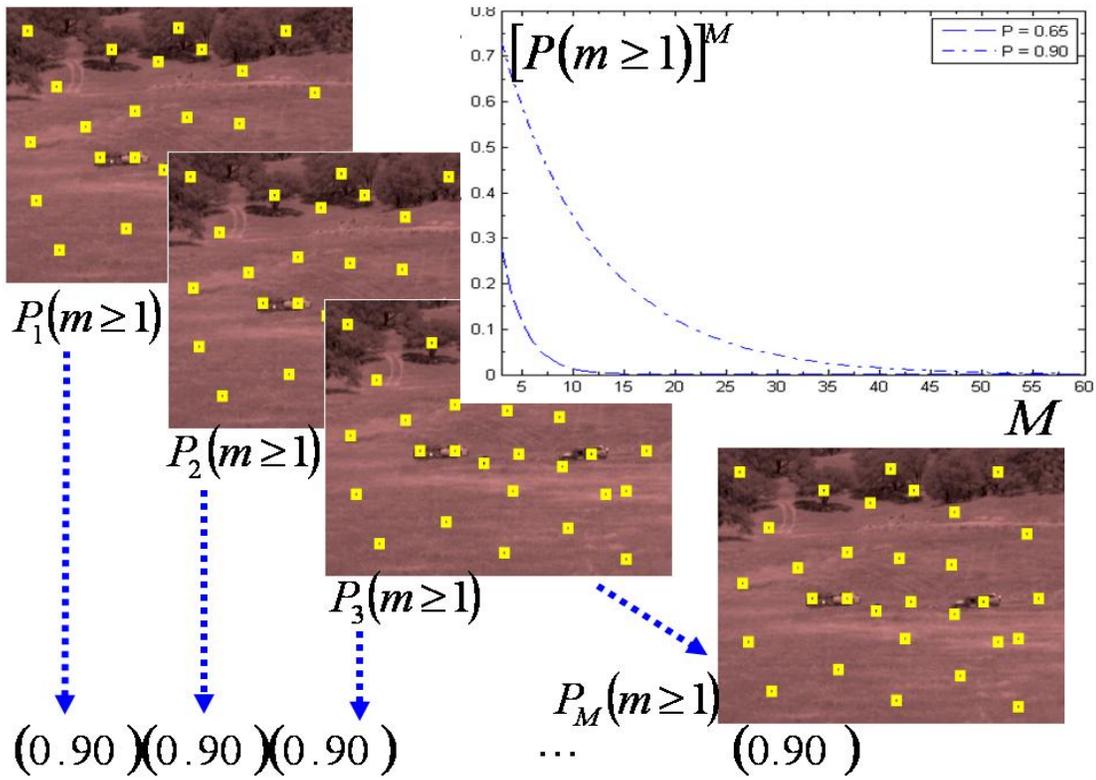
Figure 7.4. The probability $\tilde{P}(\tilde{m} = M)$ that all M random sampling processes (repetitions) will have at least a contaminated block of data decreases as a function of increasing M, given that each independent process has a probability $P_g(m \geq 1)$ of being contaminated.

Fig. 7.4 also shows a plot of $\tilde{P}$ as a function of increasing M, for $P(m \geq 1) = 0.90$ and $P(m \geq 1) = 0.65$. Taking, as an example, the $\tilde{P}$ curve in Fig. 7.4 corresponding to using $P(m \geq 1) = 0.90$ in (7.4), notice that for $M > 40$, $\tilde{P}(\tilde{m} = M)$ decreases to virtually *zero*. This outcome implies that at least one out of the $M > 40$ processes has an extremely high probability of not being contaminated, as long as, $N = 22$ and target pixels do not cover significantly more than 10% of the imagery area $(q = 0.10)$. We will show shortly in Subsection 7.3.2 how to use this autonomous

173

random sampling approach in the context of GV anomaly detection, and give some

guidelines in how to choose parameters q, N, and M. (Since the M processes are

performed independently of each other, this sampling approach will be also called the

parallel random sampling approach.)

7.3.2. GV Anomaly Detection Using No Prior Information

The GV anomaly detection problem can now be addressed using (i) the

parallel random sampling approach discussed in Subsection 7.3.2 (needed to

characterize the unknown clutter background in the imagery), (ii) an effective

anomaly detector to test reference data against the entire imagery, (iii) a way to fuse

the results from testing N randomly chosen blocks of data against the entire imagery

using small windows (this will produce a 2-dim output surface per process), and (iv) a

way to fuse M independently produced 2-dim output surfaces into a single 2-dim

decision surface.

We start by choosing from Chapter 4 a multivariate detector (RX) and a univariate

detector (AVT), and follow with a discussion on how to approach (iii) and (iv) using

(i) with these detectors.

Let a GV HS data $(R \times C \times K)$ cube $\mathbf{X}$, see (2.4), be available for autonomous

testing. Let also N blocks $(n \times n)$ of data be randomly selected from the $\mathbf{X}$'s $R \times C$

spatial area and used as a reference library set $\mathbf{W}_2^{(f)}$ $(f = 1,2,\cdots,N)$ representing

clutter background spectra, where $\mathbf{W}_2^{(f)} = \left(\mathbf{y}_{21}^{(f)}, \cdots, \mathbf{y}_{2n_2}^{(f)}\right)$ is a rearranged sequence

version of the fth block of data having $n_2 = n^2$ spectra, where $\left\{\mathbf{y}_{2u}^{(f)}\right\}_{u=1}^{n_2} \in \mathbf{R}^K$ are K-

dim column vectors. Let $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \mathbf{y}_{1n_1}\right)$ be the rearranged version of a $(n \times n)$

window of test data at location ij in $\mathbf{X}$ —see (2.8) for column vectors $\left\{ \mathbf{y}_{1h} \right\}_{h=1}^{n_1} \in \mathbf{R}^K$;

first, we would like to automatically test $\mathbf{W}_1$ against all $\left\{ \mathbf{W}_2^{(f)} \right\}_{f=1}^{N}$, and produce a

single output (scalar) value $\tilde{Z}_{RX}^{(ij)} \geq 0.0$ from these N test results. Using in this case the

RX detector, see (4.8), as the base detector, we propose the following:

$$\tilde{Z}_{RX}^{(ij)} = \min_{1 \leq f \leq N} Z_{RX}^{(ij)(f)}, \tag{7.5}$$

where

$$Z_{RX}^{(ij)(f)} = \frac{n_1 n_2}{n_1 + n_2} \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2^{(f)} \right)^t \left(\hat{\boldsymbol{\Sigma}}_2^{(f)} \right)^{-1} \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2^{(f)} \right), \tag{7.6}$$

$\left\{ Z_{RX}^{(ij)(f)} \right\}_{f=1}^{N} \geq 0.0$, $\overline{\mathbf{y}}_1 = n_1^{-1} \sum_{u=1}^{n_1} \mathbf{y}_{1u}$, $\overline{\mathbf{y}}_2^{(f)} = n_2^{-1} \sum_{u=1}^{n_2} \mathbf{y}_{2u}^{(f)}$,

$\hat{\boldsymbol{\Sigma}}_2^{(f)} = \left(n_2 - 1 \right)^{-1} \sum_{u=1}^{n_2} \left(\mathbf{y}_{2u}^{(f)} - \overline{\mathbf{y}}_2^{(f)} \right) \left(\mathbf{y}_{2u}^{(f)} - \overline{\mathbf{y}}_2^{(f)} \right)^t$, $n_1 = n_2 = n^2$, and $\left(i = 1, \cdots, R - n - 1 \right)$ and

$\left(j = 1, \cdots, C - n - 1 \right)$ index the left upper corner pixel of an $n \times n$ window in $\mathbf{X}$ [see,

for instance, (2.8) and (2.9)].

Notice that if $Z_{RX}^{(ij)(1)}, Z_{RX}^{(ij)(2)}, \cdots, Z_{RX}^{(ij)(N)}$ are placed in ascending order and

denoted by $Z_{RX(1)}^{(ij)}, Z_{RX(2)}^{(ij)}, \cdots, Z_{RX(N)}^{(ij)}$, such that $Z_{RX(1)}^{(ij)} \leq Z_{RX(2)}^{(ij)} \leq \cdots \leq Z_{RX(N)}^{(ij)}$, then

$\tilde{Z}_{RX}^{(ij)} = Z_{RX(1)}^{(ij)}$ —the lowest order statistics [25].

Notice also that if $\mathbf{W}_1$ is significantly different from all $\left\{ \mathbf{W}_2^{(f)} \right\}_{f=1}^{N}$, then all of

the corresponding results $\left\{ Z_{RX}^{(ij)(f)} \right\}_{f=1}^{N}$ in (7.6) would yield high values; this outcome

means that the lowest order statistics $\tilde{Z}_{RX}^{(ij)}$ in (7.5) would also produce a high value.

Otherwise, if $\mathbf{W}_1$ is significantly similar to at least one of the samples in $\left\{\mathbf{W}_2^{(f)}\right\}_{f=1}^N$,

then at least one of the corresponding results in $\left\{Z_{RX}^{(ij)(f)}\right\}_{f=1}^N$ would yield a low value;

this low value would be assigned to $\tilde{Z}_{RX}^{(ij)}$, according to (7.5).

Since it is unknown a priori whether target spectra are present in $\mathbf{X}$, the entire

$\mathbf{X}$ needs to be tested. In order to do it, all $\left\{\tilde{Z}_{RX}^{(ij)}\right\}_{i=1,\,j=1}^{R-n-1,\,C-n-1}$ must be computed

according to (7.5), producing a 2-dim output surface $\tilde{\mathbf{Z}}_{RX}^{(g)}$, or

$$
\tilde{\mathbf{Z}}_{RX}^{(g)} = \begin{bmatrix}
\tilde{Z}_{RX}^{(11)(g)}, & \tilde{Z}_{RX}^{(12)(g)}, & \cdots, & \tilde{Z}_{RX}^{[1(C-n-1)](g)} \\
\tilde{Z}_{RX}^{(21)(g)}, & \tilde{Z}_{RX}^{(22)(g)}, & \cdots, & \tilde{Z}_{RX}^{[2(C-n-1)](g)} \\
\vdots & \vdots & \vdots & \vdots \\
\tilde{Z}_{RX}^{[(R-n-1)1](g)}, & \tilde{Z}_{RX}^{[(R-n-1)2](g)}, & \cdots, & \tilde{Z}_{RX}^{[(R-n-1)(C-n-1)](g)}
\end{bmatrix}, \qquad (7.7)
$$

where the index g $(1 \le g \le M)$ has been introduced to results produced by (7.5) in

order to denote the repetition (or process) number discussed in Subsection 7.3.1.

(Notice that $\tilde{\mathbf{Z}}_{RX}^{(g)} \in \mathbf{R}^{(R-n-1)\times(C-n-1)}$, which for $n > 1$ is a smaller spatial area than the

$\mathbf{X}$'s $R \times C$ spatial area.)

The result in (7.7) is our approach to (iii), see the first paragraph in this

subsection.

The procedure discussed thus far in this subsection will be independently

repeated M number of times, as discussed in Subsection 7.3.1. Using the pixel values

$\widetilde{Z}_{RX}^{(ij)(g)}$ from $\widetilde{\mathbf{Z}}_{RX}^{(g)}$, our approach to (iv) is to sum M results as follows: (the rationale will be explained shortly)

$$
\mathbf{Z}_{RX} = \begin{bmatrix}
\sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{(11)(g)}, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{(12)(g)}, & \cdots, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{[1(C-n-1)](g)} \\
\sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{(21)(g)}, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{(22)(g)}, & \cdots, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{[2(C-n-1)](g)} \\
\vdots & \vdots & \vdots & \vdots \\
\sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{[(R-n-1)1](g)}, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{[(R-n-1)2](g)}, & \cdots, & \sum\limits_{g=1}^{M} \widetilde{Z}_{RX}^{[(R-n-1)(C-n-1)](g)}
\end{bmatrix}. \qquad (7.8)
$$

Fig. 7.5 illustrates $\widetilde{\mathbf{Z}}_{RX}^{(g)}$ (7.7) and $\mathbf{Z}_{RX}$ (7.8) through a parallel random sampling diagram. The diagram shows M independent (parallel) paths, where, in each path, independent blocks of data are randomly selected from the input HS data cube so that the entire data cube can be tested, against these blocks of data, using a testing window of the same block size. Each path, which is indexed by g $(1 \le g \le M)$, produces a 2-dim output surface $\left(\widetilde{\mathbf{Z}}_{RX}^{(g)}\right)$, where, at the backend of the diagram, all $\left\{\widetilde{\mathbf{Z}}_{RX}^{(g)}\right\}_{g=1}^{M}$ are summed *pixelwise* (i.e., only the pixel values at the same pixel location are added), producing a final 2-dim surface $\mathbf{Z}_{RX}$, as shown in (7.8).
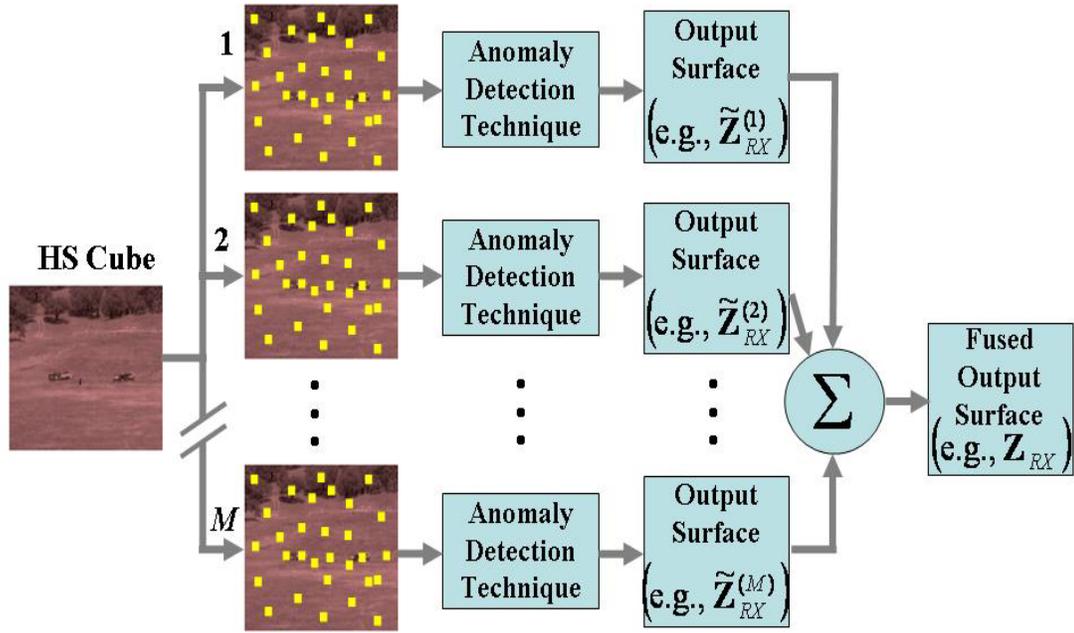
Figure 7.5. Parallel random sampling approach for GV anomaly detection, where detector's output surfaces are fused by summing pixelwise the surfaces.

For a given repetition g $(1 \leq g \leq M)$, assume that the realization of $\mathbf{W}_1$ from a window location ij in $\mathbf{X}$ is a spectral sample of a target, and the realizations of $\left\{ \mathbf{W}_2^{(f)} \right\}_{f=1}^N$ are samples of various materials composing the clutter background in $\mathbf{X}$, i.e., the randomly selected blocks of data are not contaminated with target spectra. Using an effective anomaly detector, (7.5) is expected to yield a high value for that ij location. Moreover, if the target scale in $\mathbf{X}$ is larger than $n \times n$, then the target will be represented by multiple pixels in $\tilde{\mathbf{Z}}_{RX}^{(g)}$—see (7.7), having high values. These pixels are expected to be clustered, hence, accentuating the target spatial location in $\tilde{\mathbf{Z}}_{RX}^{(g)}$. However, as discussed in Subsection 7.3.1, the contamination probability $P(m \geq 1)$, for a given g, increases as a function of increasing N, see Fig. 7.3. Fig. 7.4 shows

178

further that for a fixed q, N and an adequately large M, if (for instance) results

$\tilde{Z}_{RX}^{(22)(1)}, \tilde{Z}_{RX}^{(22)(2)}, \cdots, \tilde{Z}_{RX}^{(22)(M)}$ correspond to the same portion of the target at testing

window location $(i = 2, j = 2)$, then (7.4) give us the confidence that at least one term

in $\tilde{Z}_{RX}^{(22)(1)}, \tilde{Z}_{RX}^{(22)(2)}, \cdots, \tilde{Z}_{RX}^{(22)(M)}$ will have a high value with high probability

$\left[1.0 - \tilde{P}(\tilde{m} = M)\right]$; we can capture this high value(s) by summing these terms, or for

this example $\sum_{g=1}^{M} \tilde{Z}_{RX}^{(22)(g)}$, as shown in (7.8) for all ij locations. Notice that a target may

also be represented by multiple (clustered) pixel locations in $\mathbf{Z}_{RX}$ (7.8).

The implementation described in this subsection for the RX detector is readily

applicable to other multivariate or univariate detectors, described in Chapter 4, by

merely using $\mathbf{W}_2^{(f)}$ in place of $\mathbf{W}_2$, and applying the corresponding formulas

accordingly.

Next, we discuss implementation of the GV anomaly detection approach using the

univariate AVT detector as base detector.

To use the univariate AVT detector, spectral samples must be first

transformed using the transformation method discussed in Subsection 3.2. We can do

that by using the N randomly selected blocks of data and arrange to

$\mathbf{W}_2^{(f)} \left(1 \le f \le N\right)$, replacing $\mathbf{W}_2$ in (3.1) with $\mathbf{W}_2^{(f)}$, and using the index f,

accordingly, in (3.3), (3.5), and (3.7), or

$$\mathbf{W}_2^{(f)} = \begin{bmatrix} \mathbf{y}_{21}^{(f)}, & \cdots, & \mathbf{y}_{2n_2}^{(f)} \end{bmatrix}$$

$$= \begin{bmatrix} L_{211}^{(f)}, & \cdots, L_{21n_2}^{(f)} \\ L_{221}^{(f)}, & \cdots, L_{22n_2}^{(f)} \\ L_{231}^{(f)}, & \cdots, L_{23n_2}^{(f)} \\ & \vdots \\ L_{2(K-2)1}^{(f)}, & \cdots, L_{2(K-2)n_2}^{(f)} \\ L_{2(K-1)1}^{(f)}, & \cdots, L_{2(K-1)n_2}^{(f)} \\ L_{2K1}^{(f)}, & \cdots, L_{2Kn_2}^{(f)} \end{bmatrix}, \tag{7.9}$$

$$\nabla_2^{(f)} = \begin{bmatrix} \left(L_{221}^{(f)} - L_{211}^{(f)} \right), & \cdots & , \left(L_{22n_2}^{(f)} - L_{21n_2}^{(f)} \right) \\ \left(L_{231}^{(f)} - L_{221}^{(f)} \right), & \cdots & , \left(L_{23n_2}^{(f)} - L_{22n_2}^{(f)} \right) \\ & \vdots \\ \left(L_{2(K-1)1}^{(f)} - L_{2(K-2)1}^{(f)} \right), & \cdots, & \left(L_{2(K-1)n_2}^{(f)} - L_{2(K-2)n_2}^{(f)} \right) \\ \left(L_{2K1}^{(f)} - L_{2(K-1)1}^{(f)} \right), & \cdots, & \left(L_{2Kn_2}^{(f)} - L_{2(K-1)n_2}^{(f)} \right) \end{bmatrix}, \tag{7.10}$$

$$\overline{\nabla}_2^{(f)} = \frac{1}{n_2} \nabla_2^{(f)} \mathbf{1}_{n_2 \times 1} \tag{7.11}$$

and, denoting the columns of $\nabla_2^{(f)}$ as $\left\{ \nabla_{2u}^{(f)} \right\}_{u=1}^{n_2}$,

$$\left\{ x_{2u}^{(f)} = \frac{180}{\pi} \arccos \left(\frac{\left(\nabla_{2u}^{(f)} \right)^t \overline{\nabla}_2^{(f)}}{\left\| \nabla_{2u}^{(f)} \right\| \left\| \overline{\Delta}_2^{(f)} \right\|} \right) \right\}_{u=1}^{n_2}. \tag{7.12}$$

And equivalently for $\mathbf{W}_1 = \left(\mathbf{y}_{11}, \cdots, \ \mathbf{y}_{1n_1} \right)$—the rearranged version of a $\left(n \times n \right)$ window of test data at location ij in $\mathbf{X}$, using (3.2), (3.4), (3.6), and the columns of $\nabla_2^{(f)}$ in (7.10)—$\left\{ \nabla_{2u}^{(f)} \right\}_{u=1}^{n_2}$, we have

$$\left\{ x_{1u}^{(f)} = \frac{180}{\pi} \arccos\left(\frac{\left(\nabla_{2u}^{(f)} \right)^t \overline{\nabla}_1}{\left\| \nabla_{2u}^{(f)} \right\| \left\| \overline{\Delta}_1 \right\|} \right) \right\}_{u=1}^{n_2}. \tag{7.13}$$

From (7.12) and (7.13), the following two univariate sequences will be used as inputs to the AVT detector:

$$x_2^{(f)} = \left(x_{21}^{(f)}, \ x_{22}^{(f)}, \ \cdots, \ x_{2n_2}^{(f)} \right) \tag{7.14}$$

and

$$x_1^{(f)} = \left(x_{11}^{(f)}, \ x_{12}^{(f)}, \ \cdots, \ x_{1n_2}^{(f)} \right), \tag{7.15}$$

where $1 \leq f \leq N$.

Following the discussion that led to (7.5), the AVT detector—see (4.41)—is implemented as follows:

$$\tilde{Z}_{AVT}^{(ij)} = \min_{1 \leq f \leq N} Z_{AVT}^{(ij)(f)}, \tag{7.16}$$

where,

$$Z_{AVT}^{(ij)(f)} = n_2 \frac{\left(S_2^{2(f)} - S_{union}^{2(f)}\right)^2}{\hat{\varsigma}_2^{2(f)}}, \qquad (7.17)$$

S_2^2 is the sample variance of $x_2^{(f)}$ in (7.14), $S_{union}^{2(f)}$ is the sample variance of $\left(x_2^{(f)}, x_1^{(f)}\right)$ —the combined sample using (7.14) and (7.15), and—for $\bar{x}_2^{(f)}$ denoting the sample mean of $x_2^{(f)}$—

$$\hat{\varsigma}_2^{2(f)} = \sum_{u=1}^{n_2} \frac{\left[\left(x_{2u}^{(f)} - \bar{x}_2^{(f)}\right)^2 - S_2^{2(f)}\right]^2}{n_2 - 1}. \qquad (7.18)$$

After computing all $\left\{\widetilde{Z}_{AVT}^{(ij)}\right\}_{i=1,\ j=1}^{R-n-1,\ C-n-1}$ using (7.16) and indexing them with the given repetition g $(1 \le g \le M)$, a 2-dim output surface $\widetilde{\mathbf{Z}}_{AVT}^{(g)}$ is produced,

$$\widetilde{\mathbf{Z}}_{AVT}^{(g)} = \begin{bmatrix} \widetilde{Z}_{AVT}^{(11)(g)}, & \cdots, & \widetilde{Z}_{AVT}^{[1(C-n-1)](g)} \\ \vdots & \vdots & \vdots \\ \widetilde{Z}_{AVT}^{[(R-n-1)1](g)}, & \cdots, & \widetilde{Z}_{AVT}^{[(R-n-1)(C-n-1)](g)} \end{bmatrix}, \qquad (7.19)$$

which leads to AVT's final output surface $\mathbf{Z}_{AVT}$,

$$\mathbf{Z}_{AVT} = \begin{bmatrix} \sum_{g=1}^{M} \widetilde{Z}_{AVT}^{(11)(g)}, & \cdots, & \sum_{g=1}^{M} \widetilde{Z}_{AVT}^{[1(C-n-1)](g)} \\ \vdots & \vdots & \vdots \\ \sum_{g=1}^{M} \widetilde{Z}_{AVT}^{[(R-n-1)1](g)}, & \cdots, & \sum_{g=1}^{M} \widetilde{Z}_{AVT}^{[(R-n-1)(C-n-1)](g)} \end{bmatrix}. \tag{7.20}$$

For the remainder of this chapter, we will refer to parallel random sampling as PRS and, consequently, to (7.8) and (7.20) as PRS-RX and PRS-AVT, respectively.

7.3.3. Summary of Results

This subsection focuses on the application of the PRS approach, discussed in Subsection 7.3.2, to the autonomous GV anomaly detection problem. No prior information (e.g., spectral library, expected target scales, any knowledge about the scenario) is used, except for the comparative analysis discussed later in this chapter. Since this approach requires an effective anomaly detector as its base detector, and results from Chapter 5 and Chapter 6 showed that the two-step univariate detection techniques are more effective testing difficult simulated cases than existing multivariate detection techniques, most of the results presented herein were obtained using PRS-AVT. Initial results using PRS-AVT and PRS-RX are shown in Subsection 7.3.3.1; additional results are shown in Subsection 7.3.3.2 applying PRS-AVT to data collected recently (May/June 2008) at Picatiny Arsenal (New Jersey), exemplifying various scene conditions (e.g., fog, partially overcast); and finally, using prior information (manual sampling of the background clutter), additional comparative results are presented in Subsection 7.3.3.3 using multiple multivariate and univariate detectors, including RX and AVT. Subsection 7.3.3.1 focuses on first

checking whether PRS (using an effective detector) works; Subsection 7.3.3.2 focuses

on the performance robustness of PRS-AVT to changing illumination environment

and atmospheric conditions; and Subsection 7.3.3.3 focuses on a comparative analysis

using the different detection techniques discussed in Chapter 4 for the GV anomaly

detection problem.

7.3.3.1.   Initial Results Using No Prior Information

PRS-AVT was initially applied to Cube 1, Cube 2, Cube 3, and Cube 4

(see Fig. 7.1) to test for scene (spectral) anomalies, obtaining excellent results—they

are shown in this subsection. PRS-RX was also applied to Cube 3, since this cube was

used for the data characterization discussed in Section 2.4.

We begin by first showing how parameters N and M affect the output of

PRS-AVT testing Cube 1, see Fig. 7.6. Fig 7.6 (top right and bottom left) represent

two different outcomes for $\mathbf{Z}_{AVT}$ in (7.20), where $n \times n$ was fixed at once to

20×20 (for all data blocks and window sizes) and parameters q, N, and M were set

to $(q = 0.1; N = 3; M = 3)$—top right display—and

$(q = 0.1; N = 22; M = 40)$—bottom left display. (These output surfaces, which

for displaying purposes were extended to the size of Cube 1, are displayed using a

pseudocolor map, such that, the brighter the pixel values in those surfaces, the

stronger it is the evidence of anomalies at those pixel locations, relative to randomly

selected blocks of data. Also, for calibration purposes, the single motor vehicle at the

scene's center right has about 25,000 pixels, which means that a 20×20 window

would test 63 non-overlapping pixel locations over that target; but this target would

yield significantly more than 63 pixels in the output surface since sliding windows overlap.) The top right output surface displays an example when N is not set sufficiently high in order to adequately represent the clutter background. In this case, three blocks of data were randomly selected from the scene (most likely from the open field area, since it is the largest area in the scene), and used by the AVT detector to suppress [according to $\tilde{\mathbf{Z}}_{AVT}^{(g)}\left(g=1,\cdots,M\right)$ in (7.19)] the open field in Cube 1, not only once, but most likely $M=3$ times. As a result, the three motor vehicles and the canopy area on the upper portion of that scene were accentuated relative to the open field. Initially, we ignored the Binomial distribution model and set parameters N and M intentionally low in order to test Cube 1 and show the undesired result in Fig. 7.6 (top right).

If M were set much higher (e.g., 30), with other parameters fixed, one or more $\tilde{\mathbf{Z}}_{AVT}^{(g)}$ would most likely have the tree area also suppressed, but since all of $\tilde{\mathbf{Z}}_{AVT}^{(g)}$ are pixelwise summed [see (7.20)] that tree area (although smaller than the open field, yet significantly larger than individual candidate targets) would still be accentuated relative to the open field. The results shown in Fig. 7.6 (top right) gave us the initial confidence that PRS seems to work as intended. We then used the Binomial distribution model to guide us on setting N and M, given a sensible q.

$$
\mathbf{Z}_{AVT} = \begin{bmatrix} \sum\limits_{g=1}^{M} \widetilde{Z}_{AVT}^{(11)(g)}, & \cdots, & \sum\limits_{g=1}^{M} \widetilde{Z}_{AVT}^{[1(C-n-1)](g)} \\ \vdots & \vdots & \vdots \\ \sum\limits_{g=1}^{M} \widetilde{Z}_{AVT}^{[(R-n-1)1](g)}, & \cdots, & \sum\limits_{g=1}^{M} \widetilde{Z}_{AVT}^{[(R-n-1)(C-n-1)](g)} \end{bmatrix}
$$
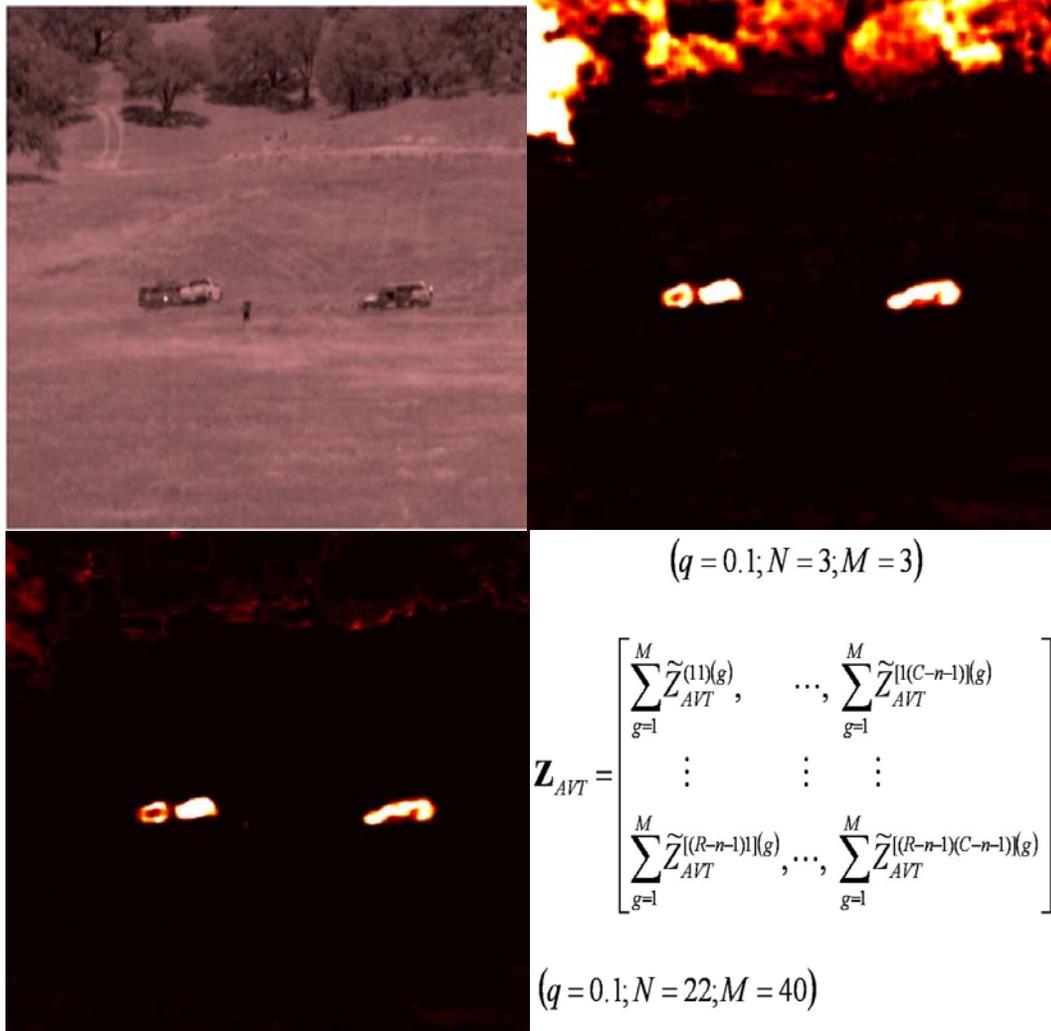
$\left(q = 0.1; N = 22; M = 40\right)$

Figure 7.6. PRS-AVT results on Cube 1 (top left) for scene anomalies; output surface (top right) using parameters $\left(q = 0.1; N = 3; M = 3\right)$; and output surface (bottom left) using parameters $\left(q = 0.1; N = 22; M = 40\right)$. Brighter pixels values in the output surfaces correspond to higher confidence on the presence of anomalies in the imagery, relative to randomly selected blocks of data. Also, notice that since $\mathbf{Z}_{AVT}$ is a sum of results, bright clusters in those surfaces are smooth clusters.

For most remote sensing applications, targets (if present in the scene) will not cover more than 10 percent of the imagery spatial area. For instance, the motor

vehicle having 25,000 pixels in Cube 1 covers 6.1% of the imagery area $\left(\dfrac{25000}{409600}\right)$.

So, we fix at once $q = 0.1$ as a robust choice. The binomial distribution plot in Fig. 7.3, for $q = 0.1$, shows that $N = 22$ yields an upper bound contamination probability $P_g(m \geq 1) \approx 0.9$ $(1 \leq g \leq M)$, and the plot in Fig. 7.4 shows a corresponding cumulative contamination probability $\tilde{P}(\tilde{m} = M) \approx 0.0$ for $M = 40$. The output surface shown in Fig 7.6 (bottom left) is the result using PRS-AVT to test Cube 1 having parameters set to $(q = 0.1; N = 22; M = 40)$. That output surface shows the manmade objects (3 motor vehicles) clearly accentuated relative to the unknown cluttered environment, given that no prior information is used about the materials composing the clutter background, or about whether targets are present in the scene, or about targets' scales relative to other object structures in the imagery. But notice in Fig. 7.6 that the standing person in the scene center is not detected, possibly because the window size might be too large and/or there must have some materials in that background (randomly selected) spectrally similar to the materials representing that person (e.g., pants, shirt, skin). Fig. 7.7 and Fig. 7.8 show additional results.

Fig. 7.7 shows results using PRS-AVT to test Cube 2 and Cube 3, and Fig. 7.8 shows results using PRS-AVT to test Cube 4, which represents a particularly difficult case of clutter suppression. Parameters were set to $(q = 0.1; N = 22; M = 40)$ for the three cubes.

Figure 7.7 PRS-AVT results on Cube 2 (top left) and Cube 3 (bottom left), where

corresponding output surfaces are shown immediately to the right of the cube displays.

Parameters were set to $(q = 0.1; N = 22; M = 40)$.

The output results shown in Fig. 7.6, Fig. 7.7, and Fig. 7.8, using

parameters set to $(q = 0.1; N = 22; M = 40)$, are excellent results for the given

application, especially for Cube 3 and Cube 4, both clearly showing the presence of a

motor vehicle highly accentuated—one in tree shades and another parked behind a

heavily cluttered environment. These results ensure to us that the idea behind the PRS

approach works, as intended, but the overall results might dependent on the effectiveness of its core anomaly detection technique. Fig 7.9 shows a qualitative comparison between using PRS-AVT and PRS-RX to test Cube 3—the HS data cube used for the data characterization discussed in Chapter 2.
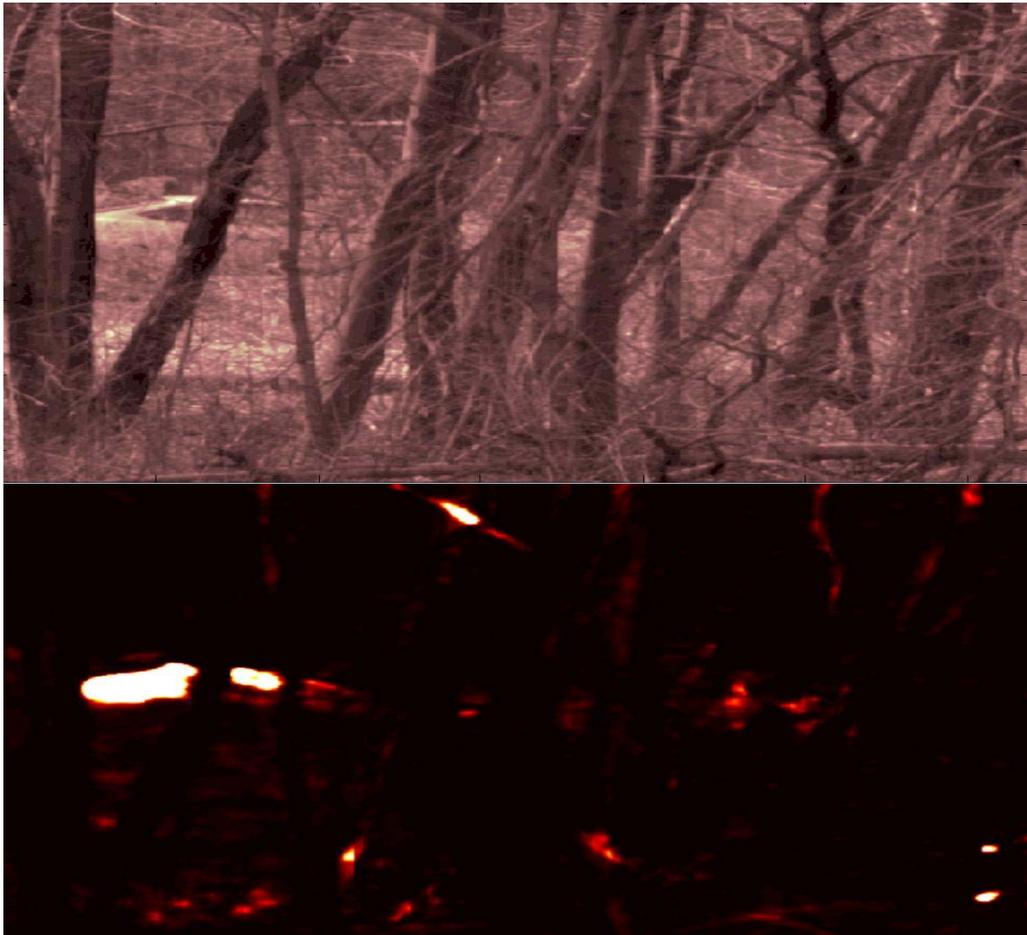
Figure 7.8. PRS-AVT results on Cube 4 (left), and corresponding output surface (right). Parameters were set to $(q = 0.1; N = 22; M = 40)$. Cube 4 exemplifies a hard case for autonomous clutter suppression.

For the results presented in Fig. 7.9, the window size $(n \times n) = (20 \times 20)$ and

parameters $(q = 0.1; M = 100)$ were fixed, but N varied $(N = 10, 50, 100)$.

Using $M = 100$ (a high number of repetitions—it took two weeks to obtain results

shown in Fig. 7.9 using the MATLAB software environment [35] and a Pentium IV

personal computer), we have confidence that the overall cumulative probabilities of

contamination for both PRS-AVT and PRS-RX are equally low for the chosen values

of N. In doing so, we can now check the sensitivity of the AVT and RX detectors to

different values of N. Under these settings, the detector that can show the lesser

sensitivity to varying N is more desired. Sensitivity can be qualitatively checked by

inspection of Fig. 7.9, i.e., the observed changes on the output surface of a given

detector as N changes. PRS-RX results are shown in the left column (Fig. 7.9), and

PRS-AVT output surfaces are shown in the right column (Fig. 7.9), where, from the

top, N values were changed from 10, 50, to 100. Both sets of output surfaces use the

same standard pseudo-color map (*rainbow*, which is available in MATLAB) for

displaying purposes. By inspection, the output surfaces in Fig 7.9 clearly show a

higher sensitivity of the RX detector to a varying N compared to the AVT detector.

These output surfaces were extended to match the approximate size of the imagery

spatial area of Cube 3 in Fig 7.9. According to available ground truth information

about the data collection, the visible clusters approximately at the center of all three

PRS-AVT output surfaces correspond to the pixel locations where a motor vehicle

happens to be present under tree shades. Similar clusters are also shown at about the

same pixel locations in PRS-RX output surfaces using $N = 50$ and $N = 100$, but

with the cost of having clusters of similar or greater strength elsewhere in the imagery
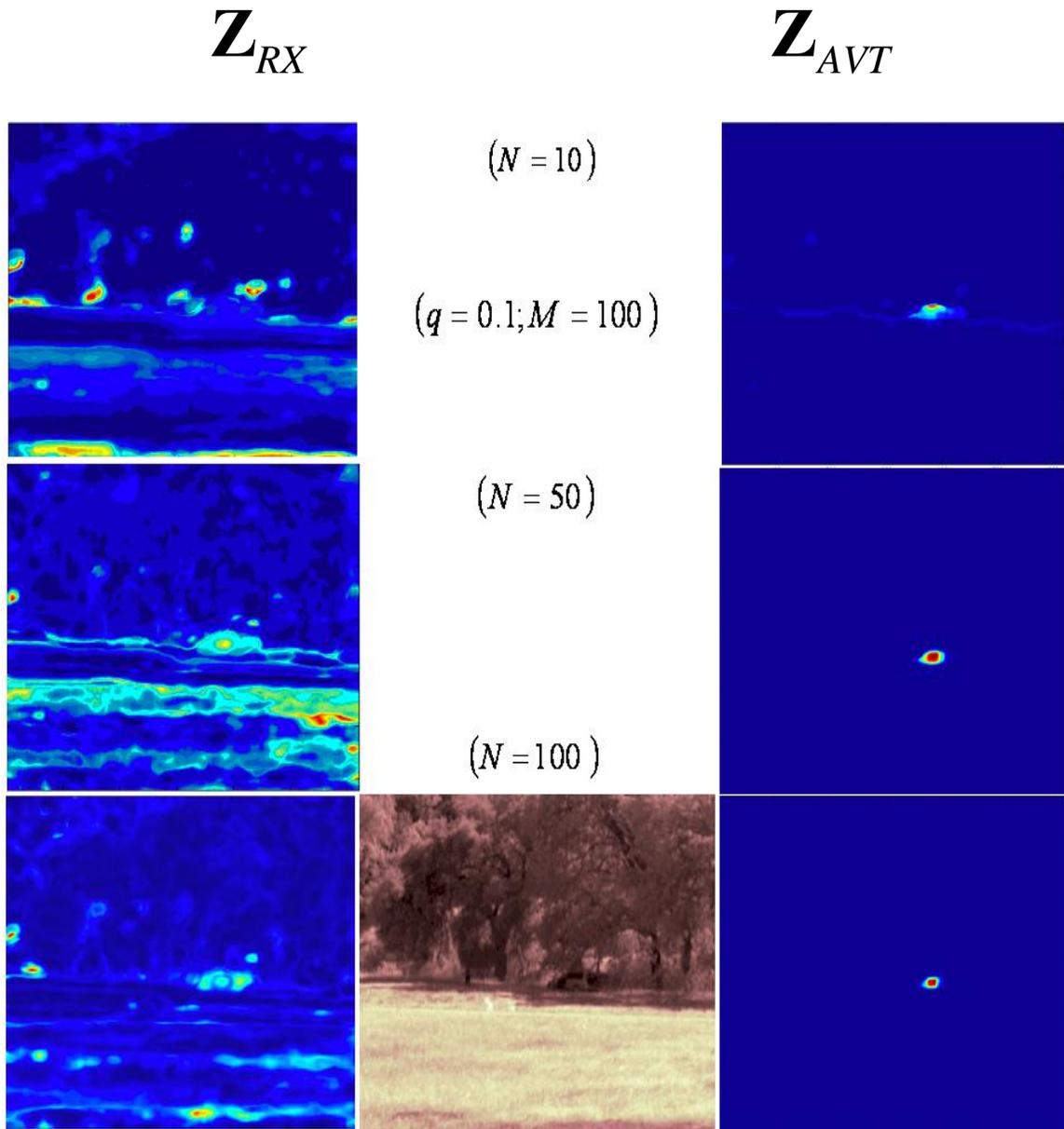
Figure 7.9. Comparison results for PRS-AVT and PRS-RX on Cube 3 (bottom center) by setting *N* to three different values (10, 50, and 100); the corresponding PRS-RX output surfaces are shown in the left column, and the corresponding PRS-AVT output surfaces are shown in the right column.

(false positives) covering some 15% to 20% of the imagery spatial area. Using $N = 10$, the corresponding PRS-RX output surface shows that the anomaly strength

at the target pixel locations are similar or greater than the strength of no more than 50% or so of all pixel locations in the imagery, which means that if a threshold is set to detect target pixels, about 50% of the imagery would show up as being anomalous to the clutter background. Many of the reasons behind the behavior of the RX detector, as shown in Fig. 7.9, were discussed in Chapter 5, and will be further discussed later in the context of the GV anomaly detection (see Subsection 7.3.3.3).

We will now address the robustness of the PRS-AVT testing HS data collected under various environmental conditions.

7.3.3.2. Adaptive Threshold Under Various Environment Conditions

The goal in this subsection is to establish an adaptive threshold method and then to test PRS-AVT for robustness using real HS imagery collected under various environmental conditions. But before we address the adaptive threshold requirement, we will first briefly introduce the additional dataset used to produce results for this subsection, followed by a brief discussion on automatically setting parameters N and M, given q.

Additional Data: Fig. 7.10 depicts photos taken at a target site under various environmental conditions at the U.S. Army Armament Research, Development and Engineering Center (ARDEC), Picatinny Arsenal, New Jersey. A mission of ARDEC is to collect data that exemplify potential challenges to candidate target detection/classification algorithms. For successful algorithms, ARDEC finds users within the Army. The target site in Fig. 7.10 features heavy clusters of trees, surrounding an open grassy field, and a dirt road leading to targets. There are two targets at that site; both are validated surrogates of military tanks, which imply that

their physical appearances are similar to the actual tanks, and they were painted using genuine paints of the targets they represent. Using the SOC-700 HS imaging system (see Fig. 2.1), HS VNIR data were recently collected (May/June 2008) from a tower standing about 0.7 km from the target site, capturing reflectance of the same site under seven different conditions: (i) clear sunny day (noon) at a higher elevation (data collected at a higher elevation angle relative to the remainder viewing perspectives in this set, except in (vii)), (ii) clear sunny day (afternoon) at a lower elevation, (iii) clear sunny late afternoon (sun light is weaker, objects cast long shadows), (iv) cloudy day (sun light energy is attenuated by some amount), (v) fog above the targets (upper fog), (vi) targets immersed in fog (lower fog), and (vii) partially overcast (where the targets are present on the overcastted portion of the scene, and elevation angle is the same as in (i)). These various conditions are known to challenge target detection/classification algorithms because they can significantly change the spectral characteristics of a particular material (e.g., paint), see, for instance, [36] and the HS sensing model discussion in Subsection 2.3—in particular the explanation for (2.1).
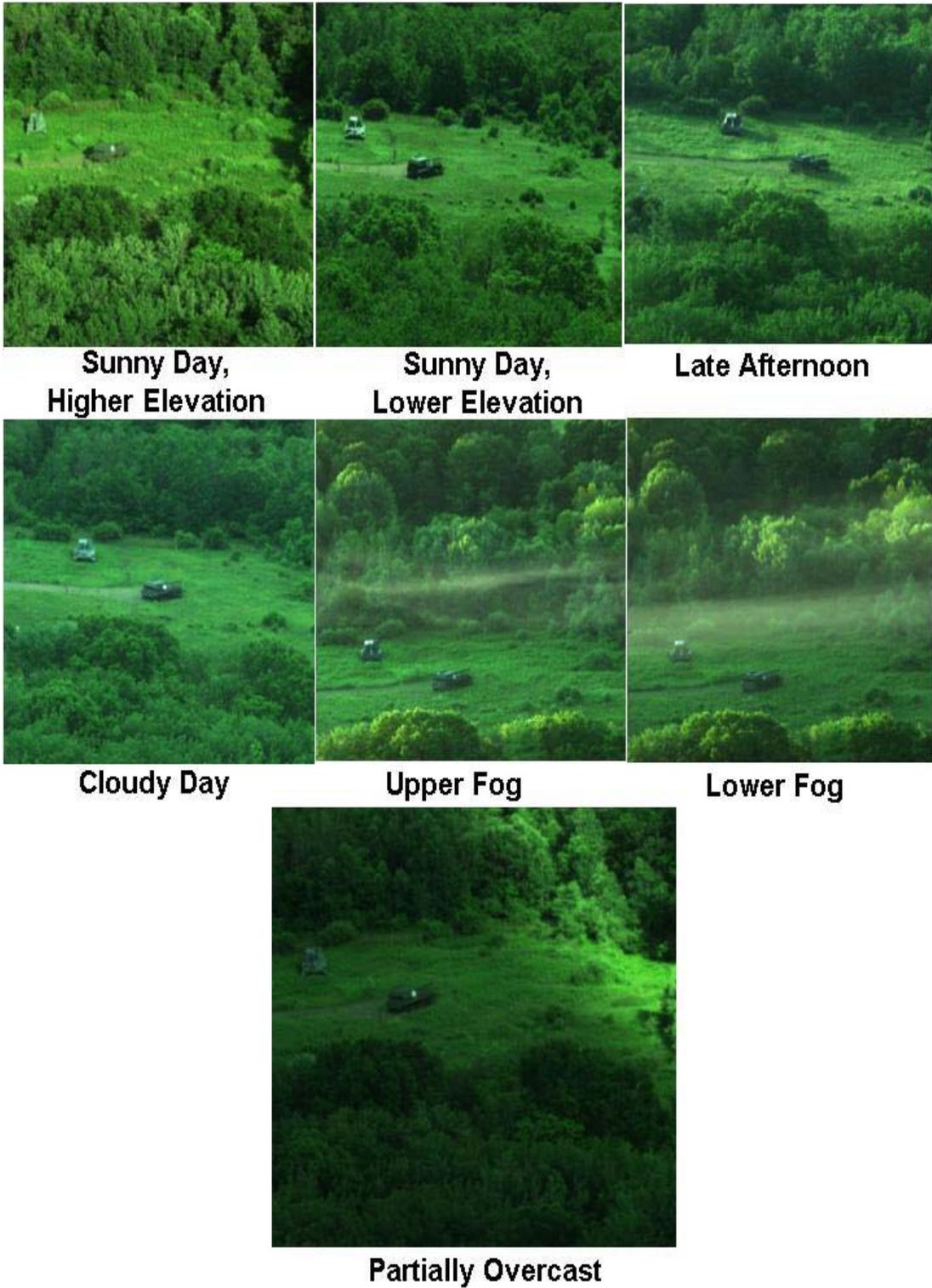
Figure 7.10. A target site under different environmental conditions.

Automatic Parameter Setting: For remote sensing applications, it is often desired to enable a machine to automatically set algorithm parameters. On this note, we can automate the setting process of parameters N and M, given q.

To properly function, the PRS approach requires an adequately large N, which undesirably increases the contamination probability $P(m \geq 1)$ per repetition, and an adequately large M, which desirably decreases the overall cumulative contamination probability $\tilde{P}(\tilde{m} = M)$ of the PRS approach for M repetitions. From (7.1), (7.2), and (7.4), using the log of base 10, direct transformation leads to

$$N = \frac{\log[1 - P(m \geq 1)]}{\log(1 - q)} \tag{7.21}$$

and

$$M = \frac{\log[\tilde{P}(\tilde{m} = M)]}{\log[1 - (1 - q)^N]}. \tag{7.22}$$

For any given q, we can fix the values of $P(m \geq 1)$ and $\tilde{P}(\tilde{m} = M)$, and obtain N and M directly using (7.21) and (7.22), respectively. As guideline, $P(m \geq 1)$ should be set high, but less than 1.0, so that N can also be relatively high and $\tilde{P}(\tilde{m} = M) < 1.0$; $\tilde{P}(\tilde{m} = M)$ should be set very low, near zero. The good news is that the actual values of $P(m \geq 1)$ and $\tilde{P}(\tilde{m} = M)$ are unimportant, as long as the guideline is followed. As an example, we could fix $P(m \geq 1) = 0.90$ and $\tilde{P}(\tilde{m} = M) = 0.01$, and

for $q = 0.05$, we obtain directly from (7.21) and (7.22) parameter values $N \approx 45$ and $M \approx 44$. (Since N and M are defined as integers, these numbers are rounded off $\approx$.) For consistency with initial results discussed in Subsection 7.3.3.1, we will fix at once $q = 0.10$, $P(m \geq 1) = 0.90$, and $P_R(m_R = M) = 0.015$, which using (7.21) and (7.22) yield $N \approx 22$ and $M \approx 40$.

Adaptive Cutoff Threshold: An adaptive cutoff threshold is also desired for remote sensing applications due to the various environmental conditions a scene can be exposed to, and to the diverse clutter background in different geographic locations across the world. For the PRS approach, we propose to take the fused output surface, in the case of AVT, $\mathbf{Z}_{AVT}$ in (7.20) and estimate both the mean and standard deviation (STD) using this surface's pixel values as input. Denoting $\mu_{\mathbf{Z}_{AVT}}$ and $\sigma_{\mathbf{Z}_{AVT}}$ the sample average and the sample STD, respectively, an adaptive cutoff threshold (a scalar) is obtained as

$$T(a) = \mu_{\mathbf{Z}_{AVT}} + a\sigma_{\mathbf{Z}_{AVT}}, \qquad (7.23)$$

where $a > 0.0$ is a constant and, using (7.19) and (7.20),

$$\mu_{\mathbf{Z}_{AVT}} = \frac{1}{(R-n-1)(C-n-1)} \sum_{i=1}^{(R-n-1)} \sum_{j=1}^{(C-n-1)} \left[\sum_{g=1}^{M} \tilde{Z}_{AVT}^{(ij)(g)} \right], \qquad (7.24)$$

and

$$\sigma_{\mathbf{Z}_{AVT}} = \sqrt{\frac{\sum_{i=1}^{(R-n-1)}\sum_{j=1}^{(C-n-1)}\left[\left(\sum_{g=1}^{M}\tilde{Z}_{AVT}^{(ij)(g)}\right) - \mu_{\mathbf{Z}_{AVT}}\right]^2}{(R-n-1)(C-n-1)-1}} \quad .$$

(7.25)

If (7.24) and (7.25) happens to be too sensitive to a relatively small number of

pixels values in $\mathbf{Z}_{AVT}$, then one could use the average median, in place of sample

mean, and a less sensitive estimate for STD. After experimenting with the latter route,

we did not see a need to follow it; hence, we chose to use (7.23), (7.24), and (7.25) to

test the additional data cubes shown in Fig. 7.10.

The SOC-700 imaging system can record the HS VNIR data cube of a site

while taking a photo of the same viewing sight of the HS imager. Fig. 7.10 depicts

only the photos of the target site. As in Subsection 7.3.3.1, the HS data cubes used for

this experiment have dimensions $R = 640$ by $C = 640$ pixels by $K = 120$ spectral

bands between 0.38 and 0.97 μm.

In order to test the additional data depicted in Fig. 7.10, we set the data block

size and testing window size to be the same, or $(n \times n) = (20 \times 20)$; $q = 0.10$;

$P(m \geq 1) = 0.90$, which using (7.21) yields $N = 22$; $P_R(m_R = M) = 0.015$, which

using (7.22) yields $M = 40$; and the adaptive cutoff threshold

$T(a) = \mu_{\mathbf{Z}_{AVT}} + a\sigma_{\mathbf{Z}_{AVT}}$, see (7.23), set initially to $T(10)$ and $T(30)$.

We tested these additional data cubes but exhibited the corresponding results

differently from the way results were exhibited in Subsection 7.3.3.1. Fig. 7.11

depicts some of those results using PRS-AVT to test the HS data cube named *Cloudy

Day*. In Fig. 7.11, the photo representing this cube is shown at the top left; the top

right shows the thresholded fused-output surface using $T(10)$; the bottom right shows the thresholded fused surface overlaid on the photo (the fused output surface and its thresholded version were automatically extended to the known photo size, as part of the overlaying process); and the bottom left shows the thresholded fused-output surface using $T(30)$. Notice in Fig. 7.11 that at 10 sigma both targets are fully detected, and the dirt road shows up as false positives. At 30 sigma, an autonomous and untrained machine—having no prior information about the target scales/shapes, or materials composing the clutter background—can detect both targets with no false alarms, Fig 7.11 (bottom left). Fig. 7.12 shows more results
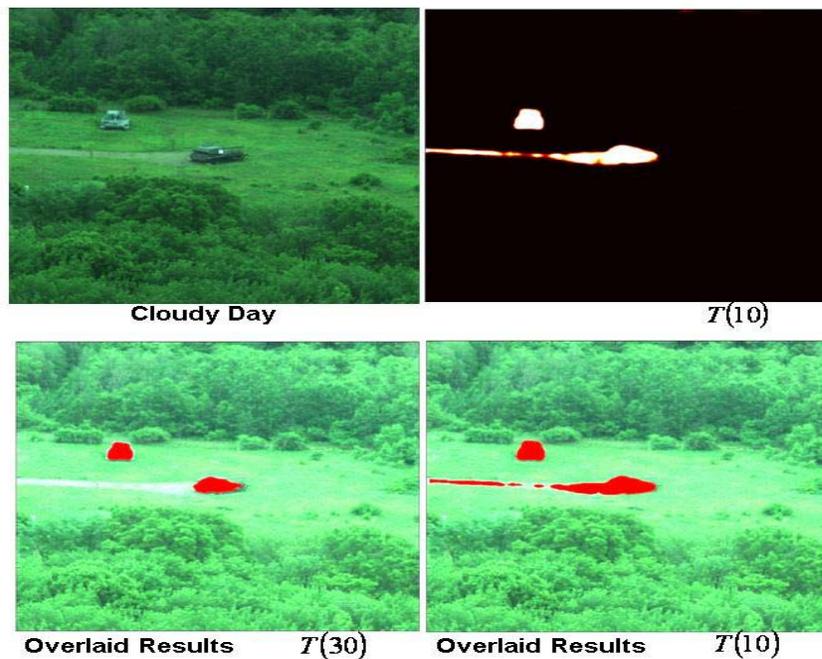
Figure 7.11. PRS-AVT thresholded fused-output surface (top right) using parameters $(q = 0.1; N = 22; M = 40)$ and $T(10)$; Overlaid results using threshold $T(10)$ (bottom right) and $T(30)$ (bottom left). At 30 sigma, both targets are fully detected with no false positives. Because of the targets' different angular orientations, they appear to have different scales and shapes.
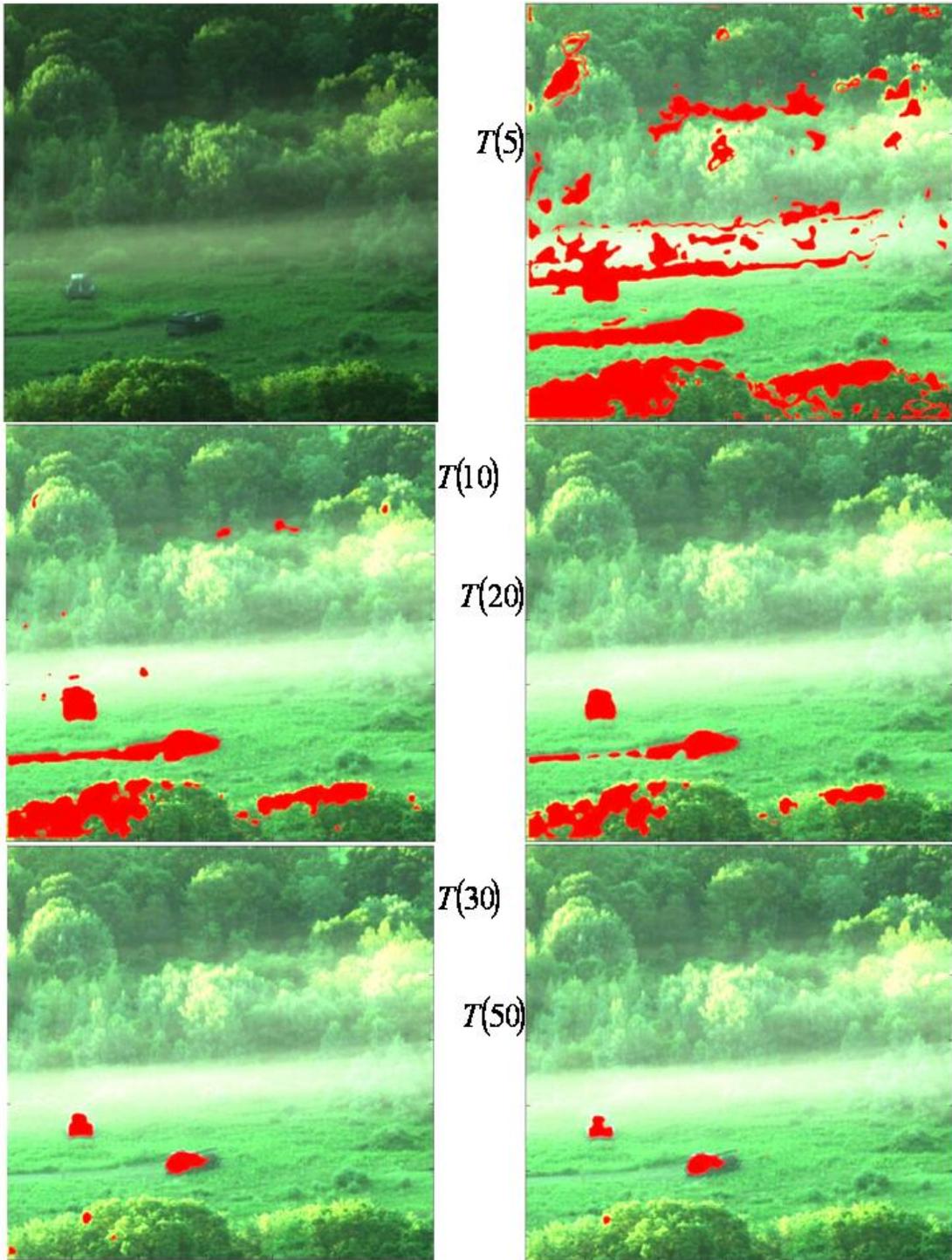
Figure 7.12. PRS-AVT overlaid resuls for Lower Fog, using parameters $(q = 0.1; N = 22; M = 40)$ and adaptive thresholds $T(5)$, $T(10)$, $T(20)$, $T(30)$, and $T(50)$.

Fig. 7.12 depicts the PRS-AVT results using the HS data cube named *Lower Fog* (arguably the most difficult one in this additional dataset for autonomous anomaly detection tasks), and applying the following thresholds: $T(5)$, $T(10)$, $T(20)$, $T(30)$, and $T(50)$, where the first column in Fig 7.12 shows results for $T(10)$ and $T(30)$, and the second column shows $T(5)$, $T(20)$, and $T(50)$. Notice that at 10 sigma, the fog over the valley causes PRS-AVT to quadruple the false alarm proportion relative to results shown in Fig. 7.11 (bottom right) for the same scene on a cloudy day. But at 30 sigma, both targets are comparably detected between HS data cubes Lower Fog and Cloudy Day, producing negligible false positives (notice in Fig 7.12, first column bottom surface, one can see very small clusters of false positives at the lower left of that surface). At 50 sigma, the detection and false alarm proportions are comparable between results using $T(30)$ and $T(50)$, which strongly suggests that PRS-AVT is capable of accentuating scene anomalies under adverse conditions. In order to check this, we tested the remainder data cubes using PRS-AVT—see results in Fig. 7.13.

In Fig. 7.13, tested cubes are shown in rows 1 and 4 (from the top), and—applying an adaptive threshold at $T(30)$—the corresponding overlaid results are shown in rows 2 and 3. Both targets are detected with virtually no false positives, except for the negligible false positive very small clusters shown on results for Lower Fog (lower left in overlaid surface). Those results suggest that the PRS approach, having an effective anomaly detection technique as its base detector, seems to be a robust approach for different data collection conditions.
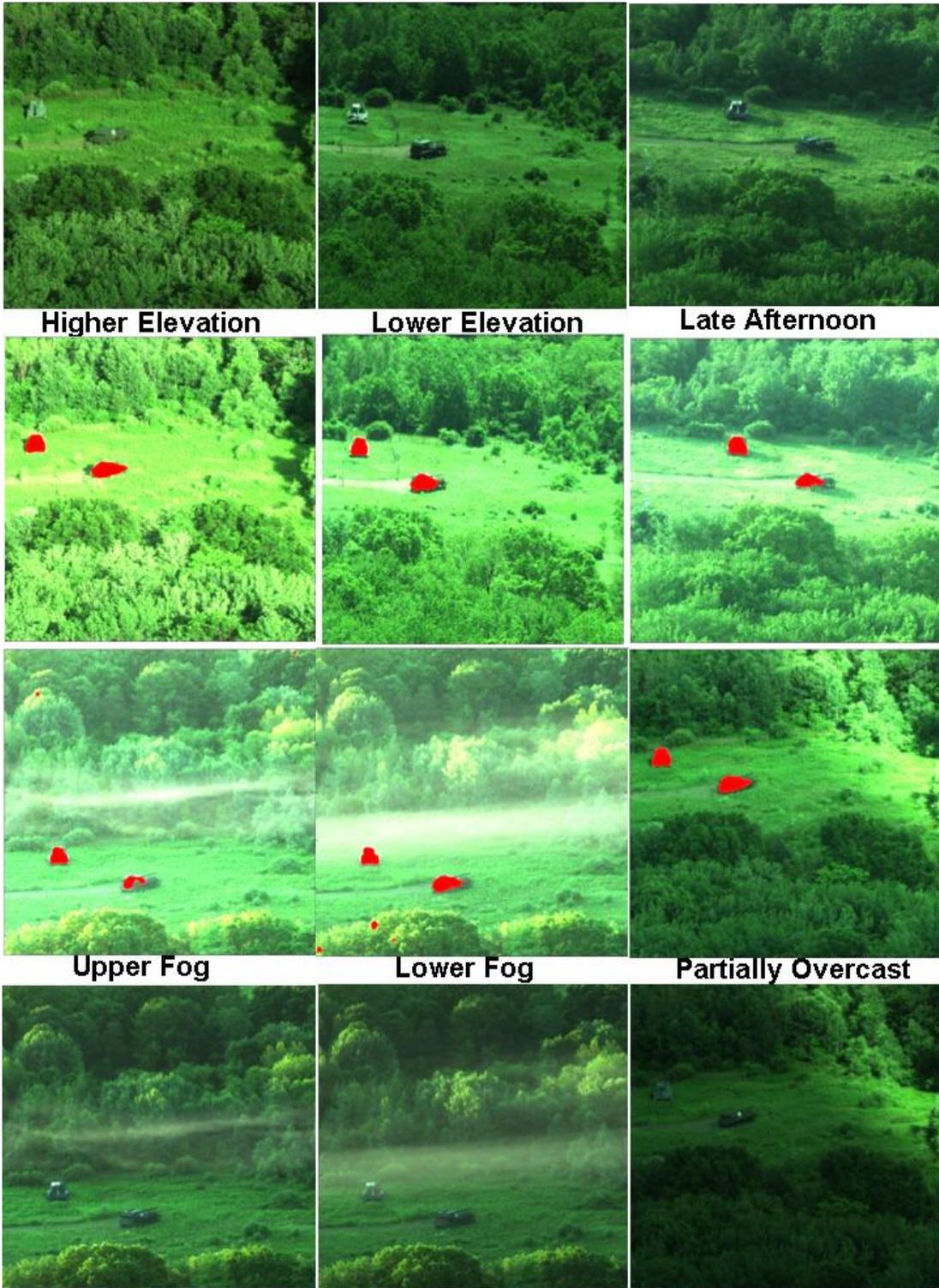
Figure 7.13. PRS-AVT overlaid results, using parameters $\left(q = 0.1; N = 22; M = 40\right)$ and adaptive threshold $T\left(30\right)$. In all cases, PRS-AVT yielded virtually zero false alarms detecting both targets.

We address next a comparative performance study among some of the multivariate and univariate anomaly detection techniques discussed in Chapter 4.

7.3.3.3.   Comparative Results Using Prior Information

For anomaly detection comparison analysis using GV imagery, the PRS approach must be decoupled from the anomaly detection technique. Although, we would like to preserve the inherent challenges of GV anomaly detection using real HS data cubes. In order to do that, we will eliminate the random sampling and the process repetition by using a man in the loop, instead, to sample spectral representatives of the clutter background from one of the imagery to be tested, i.e., differently from the discussion thus far on GV anomaly detection, prior information about the clutter background will be provided to the detectors.

The HS data cube Cube 1 was selected for background sampling by a human using prior knowledge about the imagery background (see, for instance, Fig 7.2). Two (20×20) blocks of data were selected, one representing Californian valley trees and the other representing valley terrain. These blocks of data will be the only ones used as references by anomaly detectors, as these detectors test Cube 1, Cube 2, and Cube 3 (recall that these cubes were collected from the same geo-location in California, although they represent different scenes). Proceeding thus, we will be able to check the performance of these detectors as they attempt to suppress the entire clutter background of the same cube where clutter background representatives were sampled from, and check these detectors' robustness as they attempt to suppress the clutter background of additional cubes using the limited fixed set of spectral references.

Using the notation of Subsection 7.3.2, let the rearranged spectral sample of California valley trees be denoted by $\mathbf{W}_2^{(1)}$, and the rearranged spectral sample of

| Cubes | RX | AVT |
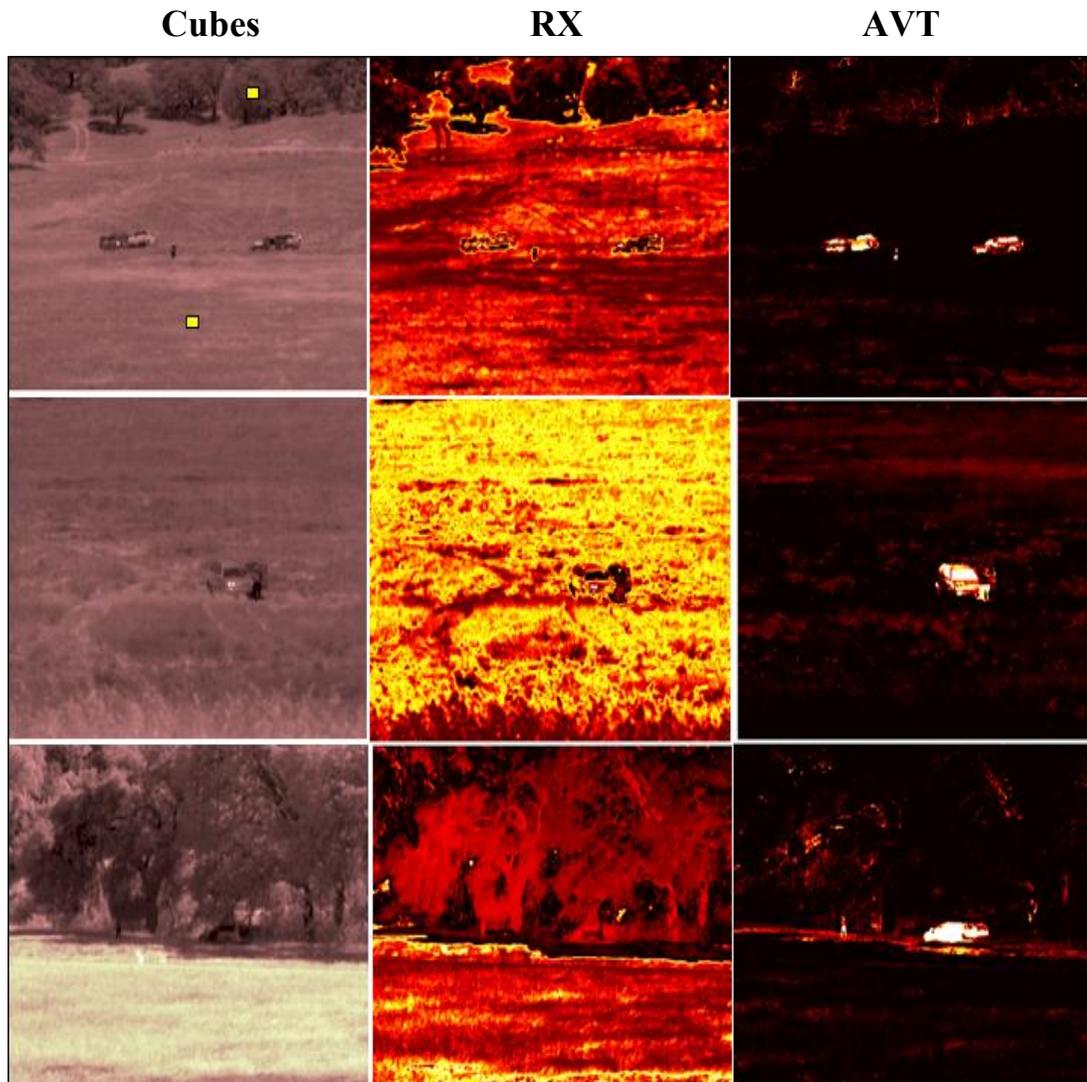
Figure 7.14. GV anomaly detection using two reference sets of spectral samples (their locations are shown as small squares in the top cube) from California trees $\mathbf{W}_2^{(1)}$ and terrain $\mathbf{W}_2^{(2)}$. The RX fused (summed) output surfaces are displayed using the same pseudo color map, where *white* depicts the strongest sign of anomalies, *yellow* strong, *red* intermediate, and *black* lowest sign of anomalies.

the California valley terrain be denoted by $\mathbf{W}_2^{(2)}$ (note that $\mathbf{W}_2^{(f)}$ $(f = 1,2)$, or $N = 2$,

for this experiment). As discussed in 8.3.2, $\mathbf{W}_1$ denotes the rearranged spectral

sample using the (20×20) test window at location ij in the spatial area of data cube

$\mathbf{X}$ being tested. For $N = 2$, the RX detector was implemented according to (7.5)

through (7.8) for this experiment, and the AVT detector was implemented using (7.9)

through (7.20). We applied the RX and the AVT detectors to Cube 1, Cube 2, and

Cube 3, and present their fused (summed) output surfaces in Fig. 7.14, columns 2 and

3, respectively.

In Fig. 7.14, we displayed all fused output surfaces using the same pseudo

color map to emphasize anomalies with respect to the reference samples by their

false-color (intensity) levels, i.e., white is equivalent to the strongest anomalies,

yellow to strong anomalies, red to intermediate anomalies, brown to weak anomalies,

black to weakest anomalies. The false colors change gradually and are relative only to

those results within the same surface, for instance, a yellow pixel in one surface does

not mean necessarily that its value is equivalent to another yellow pixel in another

surface.

The results shown in the first RX surface (row 1, column 2) are consistent

with the case studies discussed in Chapter 5. A typical multivariate detector performs

well suppressing objects in the scene having low variability and belonging to the

same material class of a reference set (e.g., the trees were suppressed). Likewise, it

performs well for accentuating objects that are significantly different from the

reference set—for instance, some parts of the vehicle at the right hand side (row 1,

column 2) were highly accentuated. (One can actually observe white pixels within the

boundaries of those vehicles by zooming close enough on both RX surfaces (rows 1 and 2, column 2), which indicates that those portions are significantly different from the reference sets.   Unfortunately, as it was observed in the top-view problem, local areas characterized by class mixtures (transition of regions) may be also accentuated by these detectors, obscuring therefore the presence of meaningful objects in that scene. In fact, for the HS cubes presented in Fig. 7.14, the RX detector seems to perform more as an edge detector than as an anomalous object detector.

On the other hand, the univariate AVT detector, which uses the data transformation discussed in Chapter 3, was able to virtually suppress the entire clutter background of Cube 1, and to accentuate large portions of the vehicles and a small portion of the standing person's pants. (Using PRS-AVT with $N \gg 2$ suppressed the distinction of the pants.) In a qualitative sense, test samples consisting of, say, a mixture of shadowed terrain and terrain were likely suppressed due to the fact that the AVT detector combines spectral samples as part of its computation, see (4.41). The reason combining samples seems to work well suppressing shadowed patches in the ground may be explained by the following: Regions characterized by tree shadows, for instance, may be interpreted as *partially obscured terrain* because tree leaves do partially obscure the incident solar light; however, since significant spectral radiances are still reflected from the partially shadowed terrain, such a region will be suppressed when compared to the union of itself and the reference set of open terrain, see (4.41).

Let us consider the results shown for Cube 2 in Fig 7.14 (row 2, columns 2 and 3). The RX surface shown in rows 2, column 2, suggests that the RX detector

205

may be susceptible to subtle spectral differences of the same terrain when observed

by the same HS sensor in a different area. Recall that Scenes 2 and 3 were tested

using the same reference sets drawn from Cube 1. The surface shown in row 2,

column 3, suggests that the AVT detector is significantly more robust to inherent

spectral differences of the same terrain.

For results testing Cube 3, see Fig 7.14 (row 3, column 2 and 3), the

interpretation of a shadowed object as a partially obscured object is especially

relevant to the interpretation of output results for Cube 3. The fused output surface

shown in Fig. 7.14, row 3, column 2, emphasizes the fact that the RX anomaly

detector performs as expected: it detects local anomalies in the scene. However, as we

have been discussing throughout the dissertation, these local anomalies are not

guaranteed to be meaningful to an image analyst in the context of the problem in

reference. For instance, in reference to the RX output surface for Cube 3, notice that

some of the tracks made by the shadowed vehicle, and the transition between the

shadowed and the non-shadowed terrain were the most anomalous regions in the

scene, as *seen* by the RX detector. The AVT detector virtually suppressed these same

regions, while the more meaningful anomalous structure (motor vehicle) was

accentuated; see the corresponding AVT surface in Fig 7.14 (row 3, column 3).

For additional comparative results, we refer to Fig. 7.15, where the

corresponding fused output surfaces are shown using the univariate detectors AVT

and AsemiP and the multivariate detectors RX, FLD, and DPC to test Cube 3.

As mentioned in Subsection 7.3.2, other multivariate detectors (e.g., FLD,

DPC) can be readily implemented by merely using $\mathbf{W}_2^{(f)}$ in place of $\mathbf{W}_2$, then

applying the corresponding formulas accordingly (see Chapter 4), and follow with the specific detector's version of (7.19) and (7.20). The FLD detector (see Subsection 4.2.2.1) and the DPC detector (see Subsection 4.2.2.2) were implemented, accordingly, for $N = 2$. Likewise, the AsemiP detector (see Subsection 4.3.1.1) was implemented, accordingly, using $\mathbf{W}_2^{(f)}$ in place of $\mathbf{W}_2$, applying the data transformation as in (7.9) through (7.15), and obtaining AsemiP's corresponding versions of (7.19) and (7.20).

Cube 3	AsemiP	AVT

RX	FLD	DPC
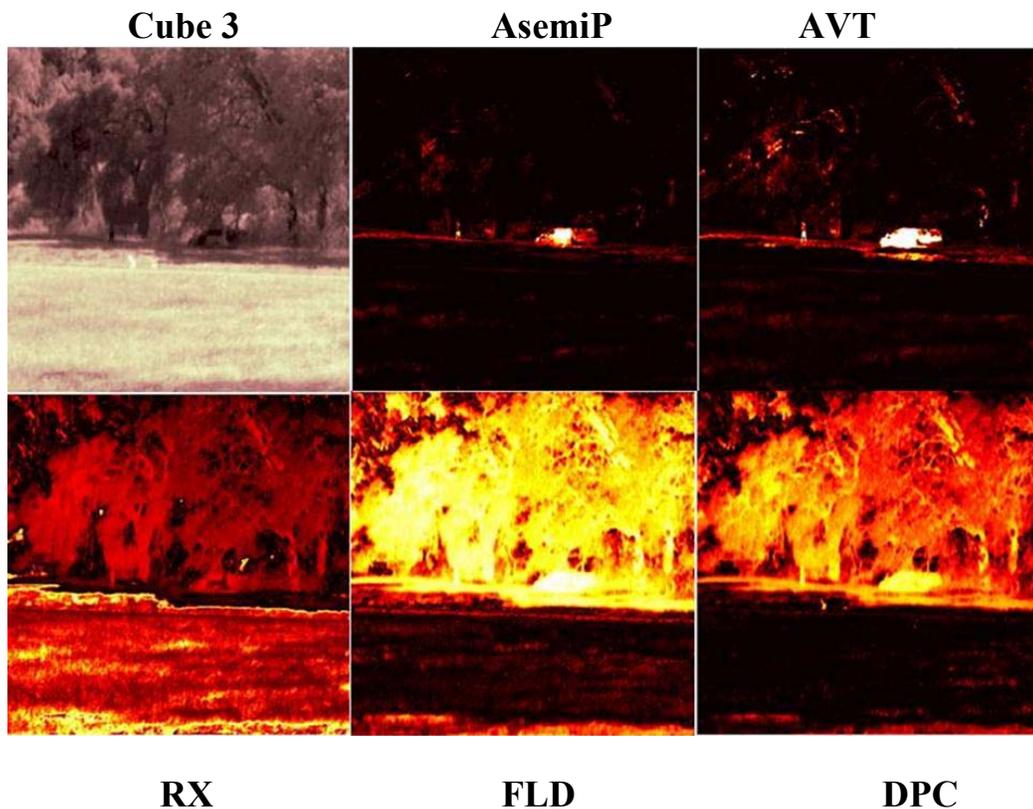
Figure 7.15 GV anomaly detection using two reference sets of spectral samples from California tree leaves and valley terrain. The reference sets were taken from Cube_1.

The RX and AVT fused output surfaces shown in Fig. 7.15 (row 2, column 1) and (row 1, column 3) are exactly the same ones corresponding to those detectors in Fig. 7.14. Notice that the FLD fused output surface shown in Fig. 7.15 (row 2, column 2) emphasizes the spectral differences between the shadowed tree region and the two reference sets, which incidentally are the same reference sets drawn from HS Cube 1. Notice that the FLD detector accentuates significantly a large portion of the shadowed motor vehicle, among other shadowed materials in that region (e.g., shadowed tree trunks and leaves). The DPC detector, on the other hand, focused on a portion of the vehicle's tire tracks as being the most anomalous object class in the entire scene relative to the reference sets $\mathbf{W}_2^{(1)}$ and $\mathbf{W}_2^{(2)}$ (the tire tracks are observed as two approximately parallel bright lines at the left of the vehicle going toward the trees, see Cube 3 in Fig. 7.15).

Taking a closer look at the DPC fused output surface in Fig. 7.15 (row 2, column 3) did reveal that about three pixels within the boundaries of the tire tracks are actually white (highest intensity). The shadowed vehicle, as well as a large portion of the shadowed tree region, produced the next lower intensity values below the tire tracks' white pixels, which indicate that a cutoff threshold would have to be set relatively low in order to detect the target in the DPC fused output surface—by inspection, it would yield about 20% of the imagery as false positives. This is comparable with the FLD detector's result, see Fig. 7.15. By inspection, in order to detect the target using the RX fuses output surface, a cutoff threshold would yield about 60% of the imagery as false positives. Both AsemiP and AVT detectors are comparable in performance testing Cube 3, and, in order to detect the target using

their fused output surfaces (see Fig. 7.15), a high cutoff threshold (e.g., 30 sigma) would yield negligible number of pixels as false positives.

The results shown in Fig. 7.14 and Fig. 7.15 suggest that the univariate anomaly detection technique, which uses the data transformation from Chapter 3 and sample-combining metrics in Chapter 4, seems more effective than conventional multivariate techniques for GV anomaly detection applications.

7.4. Summary and Conclusions

This chapter has proposed and examined the performance of an autonomous approach for the GV anomaly detection problem using real HS data cubes. The approach is generalized in the sense that it can be used with any detection technique, although this chapter showed that effectiveness of the chosen base detector will significantly affect the test results. This approach applies random sampling of the imagery and repeats the sampling process in order to mitigate the probability of contamination (spectral samples of candidate targets being used as clutter background reference samples). As such, this approach requires no prior information (e.g., a spectral library of the clutter background and/or target, target size or shape), and, therefore, is free from training requirements. This chapter showed that the PRS approach can be modeled by the binomial family of distributions, where the only target related parameter q (the upper bound proportion of target pixels potentially covering the spatial area of the imagery) is robust—thus invariant—to different sizes and shapes of targets, number of targets present in the scene, target aspect angle, partially obscured targets, or sensor viewing perspective. Binomial distribution plots were used to set other parameters: N (number of randomly selected blocks of data)

and M (number of process repetitions). This chapter also showed how N and M can be automatically set using a simple guideline, and how to implement an adaptive cutoff threshold method with PRS.

The PRS-AVT approach, in particular, was applied to real HS data cubes yielding excellent results for different target deployments (target in an open field, target in tree shades, and target behind heavy wooded region), different environmental and illumination conditions (conditions as diverse as having fog over or immersing the targets, partially overcast, different elevation angles and times of the day), and different clutter backgrounds (Californian valley, New Jersey wooded areas). Finally, a comparative analysis was presented to show the effectiveness of using the univariate anomaly detection technique, as proposed in this dissertation, over a more conventional multivariate anomaly detection technique (e.g., RX) to the GV anomaly detection problem. Chapter 6 showed similar results using TV simulated data cubes.

Chapter 8　　　Conclusions and Future Work

8.1 Summarized Conclusions

The objective of this work was to propose and evaluate specific algorithms using—as input—transformed HS data in order to obtain novel forms for output surfaces, and then use these output surfaces to improve the performance of candidate ground view and top view anomaly detection systems.

To date, a significant amount of research has focused on classification and detection algorithms using parametric HS data models as foundation for algorithm development, while little has been done to address the underlying fundamental problems that affect algorithm performances. One goal of this work was to identify a short list of fundamental performance challenges for existing local and global anomaly detection algorithms, and then use this list to find from a large population of scoring algorithms those metrics that could perform more robustly over these fundamental challenges. Another goal was to introduce to the HS research community the state of the art in global anomaly detection that would not require segmentation of the HS image data, as the state of the art in segmentation is still unreliable. (Global anomaly detection requires segmentation in the prior art—see, for instance, [12] and [13]).

This dissertation identified three underlying key factors in spectra that can interfere with detection performance, as observed through sliding small windows, and studied via simulation their effects on the performance of existing HS anomaly detection algorithms on specified null and alternative Simple hypotheses, using two-

sample data models under multivariate normal and mixtures of multivariate normal distributions. These factors are: spectral magnitude (bias), spectral shape, and spectral mixture.

A data transformation method was proposed to reduce algorithm sensitivity to spectral magnitude, while preserving high sensitivity to spectral shape; both properties are desired for effective anomaly detection, as described in Subsection 5.4.1. A semiparametric scoring metric and a few alternative scoring algorithms were proposed to handle spectral mixtures, where each scoring metric uses two transformed spectral samples.

Results from the simulation experiment study showed that different detectors fall into groups that behave differently, essentially because as algorithms they seem to be designed to pick up different features; however, as shown in tabular form in Section 5.4, it is noticeable that the two-step univariate detectors (data transformation followed by univariate scoring, as proposed in Subsection 4.3) are significantly less sensitive to spectral bias and more sensitive to spectral shape (both are desirable features) than the existing multivariate detectors chosen for the study (see Subsection 4.2.1.3 through Subsection 4.2.2.3). On average, the two-step univariate detectors also outperformed existing multivariate detectors on the challenging simulation experiments involving idealized spectral mixtures of multivariate normal distributions, see Subsection 5.4.2.

A more realistic simulation experiment was conducted to assess the performance of detectors on spectral mixtures by generating idealized multispectral (MS) data cubes, under multiple completely specified multivariate normal distributions and their

parameters. Idealized null MS data cubes were used for obtaining detectors' cutoff

thresholds, conditional to a fixed type I error for all the detectors used in the study;

and idealized alternative MS data cubes were used for obtaining detectors' power. A

key point in this simulation experiment is that the background clutter configuration

varied in complexity level from relatively easy, moderate, to hard for all the detectors

used for the study in Chapter 6, such that, an alternative MS data cube corresponding

to a given null MS data cube used the same clutter background configuration and data

specification of the given null data cube; alternative data cubes, however, featured

small blocks of data representing targets. Results tabulated in Table 6.3 through Table

6.9 suggest that the overall performance of univariate anomaly detectors can be

significantly less dependent on, or sensitive to, the background configuration of data

cubes than the overall performance of popular multivariate anomaly detectors. These

univariate detectors also outperformed the multivariate detectors on alternative data

cubes having targets in difficult background configurations (see Fig. 6.9, Table 6.9

and Table 6.10).

Finally, this dissertation presented a fully operational GV global anomaly

detection algorithm and evaluates the approach using real HS data cubes, where some

targets are present in a natural clutter background under different illumination and

atmospheric conditions. The uniqueness of this GV anomaly detection approach is

that a random sampling model was proposed as a parallel process in order to mitigate

the likelihood that samples of targets are erroneously used as clutter spectral

references during imagery testing. The cumulative probability $\tilde{P}$ of taking target

samples by chance during parallel processing were modeled by the binomial

distribution family, such that $\tilde{P} = \left[1 - (1-q)^N\right]^M$ [see, (7.4)], where N is the number of randomly selected $n \times n$ blocks of data, taken from data cube $\mathbf{X}$ (per repetition or process), M is the number of parallel processes, and q (the only target related parameter) is a proportion (an upper bound guess) of the maximum total number a of all target pixels over the total number of pixels (RC) in $\mathbf{X}$, i.e., $\left(q = \dfrac{a}{RC} \right)$. Note that q is invariant to target scale, target shape, or to the number of targets in $\mathbf{X}$. For instance, $q = 0.05$ indicates that targets in the imagery area are not expected to cover more than 5% of the entire image area in $\mathbf{X}$. Choosing a sensible q for the given application, one can use $\tilde{P} = \left[1 - (1-q)^N\right]^M$ to assist on tradeoff decisions between N and M for a desired (small) $\tilde{P}$ [e.g., $\tilde{P} = 0.01$)]. This dissertation also showed how N and M can be automatically set using a simple guideline, and how to implement an adaptive cutoff threshold method for the GV global anomaly detection algorithm suite.

The remainder of this chapter summarizes the contributions of this dissertation (Subsection 8.2), limitations (Subsection 8.3), future work (Subsection 8.4), and a brief summary (Subsection 8.5).

8.2 Contributions

The more important findings and developments of this dissertation are summarized in the following list:

- <u>The investigation of underlying fundamental challenges for HS anomaly detection:</u> Identified and studied the effect of spectral magnitude, spectral

shape, and spectral mixtures on the performance of detection algorithms; and determined that—for effective anomaly detection—an ideal detector should be sensitive to spectral shape, but insensitive to spectral magnitude, as these properties favorably affects the detector performance under unknown illumination and atmospheric conditions.

- Introduction and implementation of a data transformation method to remove detectors' sensitivity to spectral magnitude, while augmenting their sensitivity to spectral shape: It was shown via simulation that taking the radiance difference between adjacent wavelength bands—followed by angle mapping—made univariate detectors insensitive to spectral magnitude but sensitive to spectral shape, as desired.

- The first use of semiparametric algorithm for HS anomaly detection: Recognized that an univariate semiparametric scoring algorithm has a natural way of handling transitions across distinct regions (spectral mixtures) in HS image data, for sliding window based tests. Alternative univariate algorithms were also proposed to perform comparably with the semiparametric scoring metric using two-sample data as inputs.

- The first use of simulated null and alternative multivariate data cubes to analyze top-view HS anomaly detectors: Conducted innovative analysis by generating and using simulated null and alternative multispectral data cubes, through statistical modeling under multivariate normal distributions, in order to assess sliding window based HS anomaly detectors in their natural operating procedure. In was shown that existing multivariate anomaly

215

detectors are sensitive to increased background configuration complexity in the null data cubes, which increased their cutoff thresholds for the same required type I error and negatively impacted their power on the simulated alternative data cubes. On average, the two-step univariate detectors were significantly less sensitive to increasing background complexity in the null data cubes, hence, outperformed the multivariate detectors in power on the alternative data cubes.

- <u>Introduction of a novel parallel random sampling method for GV global anomaly detection applications</u>: The parallel random sampling method was modeled by the binomial distribution, and—by using the two-step univariate detection technique with this method—it was shown that parametric or nonparametric segmentation is not required, as in the prior art, to achieve effective global anomaly detection. The overall method was evaluated using real HS image data collected under various illumination and atmospheric conditions.

8.3 Limitations

The methods and techniques presented in this dissertation have the following limitations:

- <u>Anomaly Detection Is Not Target Detection</u>: A key limitation of the new top view and ground view HS anomaly detection algorithms presented herein is that the correct detection of a target does not mean that the target becomes known, but merely that the target is an anomalous object to the background clutter. In addition, correct anomaly detection may not always correspond to a correct target

216

(manmade object) detection. Natural objects, such as large isolated rocks, may be detected as candidate targets, which in a strict sense should not be regarded as a false positive.

- Sample Data and Cube Data Models: The study conducted in Chapter 5 and Chapter 6 were the first of its kind, so there were some lessons learned. In future studies there should be a randomizer varying the mixture proportions in the multivariate normal distributions for different trials in the simulation experiments presented in Chapter 5, and likewise there should be a randomizer varying the configuration of the background clutter per experimental trial in the data cubes used in the simulation experiments in Chapter 6. Proceeding as so would remove the man in the loop determining the mixture proportions in this sort of simulation experiments.

- Full Target Pixels: Although Chapter 5 showed the new anomaly detection algorithms being effective in detecting a portion of the target, targets are expected to be greater than or equal to $n \times n$, which is the area of the sliding window. Moreover, targets that are smaller than the spatial resolution of a pixel (subpixel targets) are not expected to be detected using the new algorithms, because a mixture in a pixel is quite different from a mixture of pixels representing different material types—the latter is the one addressed in this dissertation.

- Results for the VNIR Region of the Spectrum: The favorable results shown for the new GV anomaly detection algorithm on the real HS image data are limited to the different illumination and atmospheric conditions described in Chapter 7 and to the VNIR (visible to near infrared) region of the spectrum. Those results should

217

not be extrapolated for other regions of the spectrum [e.g., LWIR (longwave infrared), MWIR (midwave infrared)], because the *emissivity* property of different material types do not play a major role in the VNIR region, only the *reflectivity* property does (see, for instance, [1]).

- Data Cube Rate: A sensor to have a practical value must be able to produce a digitized representation of a scene in a rate comparable to that of a video rate (e.g., 25 to 60 frames per seconds), which is significantly above the rate of the state of the art portable HS VNIR imager. This fact would impose a major practical constraint using the new GV anomaly detection in conjunction with a HS imager for an actual surveillance task. Advances in technology, however, have been occurring in remarkable speeds since the 1990's, especially in the field of electronic technology, which make us believe that such a limitation will no longer exist in the next few years.

The concern on HS hardware speed can be also extended to the computational time required to execute the new GV anomaly detection algorithm in a computer. Algorithms that are developed to perform detection tasks using HS data are notorious for being slow (taking hours, sometimes days to operate on a data cube); not because of the algorithm itself, but because of the vast amount of data a single HS data cube actually represents. A method that is often used to reduce the computational time of HS algorithms is known in the HS research community as *spectral band selection* [3], see Section 8.4 (Future Work).

8.4 Future Work

In the future, work is needed to develop more insight into the following:

- <u>Hybrid of Detectors - Fusion</u>: A natural progression of this work is to extend the use of the anomaly detection algorithms presented in this dissertation to include a hybrid of detectors, each capturing different features of the data, in order to augment robustness. Hybrid of algorithms often requires the need of fusion techniques, using as input different output surfaces; this is an open topic of research.

- <u>Cultural Clutter Background</u>: Another natural extension of this work is to evaluate the performance of the new GV anomaly detection algorithm on the presence of particular targets (e.g., standing personnel, stationary motor vehicles) in an urban environment—cultural clutter background. In this context, it would be interesting to find out whether the autonomous random selection of blocks of data of a cultural cluttered environment (having, for instance, painted walls of buildings, sidewalks and asphalt) would have the GV algorithm performing comparably with its corresponding performance on natural clutter backgrounds. We are actively searching for such a HS dataset of cultural clutter to conduct this evaluation.

- <u>Further Evaluation of the PRS-AVT Algorithm</u>: The PRS-AVT (a GV anomaly detector) will be evaluated more extensively using additional HS datasets, and its performance will be compared to existing global anomaly detection algorithms (see, for instance, [12] and [13]). Additional evaluation will be conducted using HS image data recorded by LWIR and MWIR HS imagers.

- <u>Spectral Band Selection</u>: A goal in circumventing the speed limitation issue discussed in Section 8.3 is to use a sensor that is a compromise between hyperspectral and broadband, i.e., a sensor that collects radiance using only a few

spectral bands—a multispectral cube (e.g., 10 bands). Notice that a multispectral sensor should be able to collect data faster by an order of magnitude or two than a HS sensor can for a given swath coverage. In addition, the computational cost of detection algorithms due to this reduced amount of data representing a scene may decrease by the same order of magnitude. A key decision, however, that must be made before manufacturing multispectral sensors is to determine how many of these frequency bands are most relevant and which ones should feature in these devices. A long list of contributions can be found in the literature (see, for instance, [3]) devoted exclusively to answer this question. The conclusions of these contributions, however, independently of the method applied share explicitly, or implicitly, a common message: It depends. It depends on the specific material one is interested in detecting; it depends on the number of material types one expects to find in the same scene; it depends on the region of the spectrum the sensor is expected to operate, etc. To follow up with our research, we plan to use a favorite scoring metric (e.g., AVT), as a decision criterion, and the fact that we would like to find all types of manmade objects in different natural clutter backgrounds to determine the minimum number of combination of bands that would maximize performance on a HS dataset recorded by a particular sensor (e.g., SOC-700).

8.5 Summary

The users have an ambitious goal for target detection requirements. They would like to have an algorithm suite that can detect a large set of known targets of different sizes and shapes, at different ranges between targets and sensors, at different viewing

perspectives (top view, ground view), anywhere in the world, under unknown illumination environment and unknown atmospheric conditions.

This dissertation offers specific algorithms that could detect the presence of a large set of targets using HS image data, while satisfying many of the users' requirements; however, targets would not be detected as specific manmade objects (which are often composed of multiple material types), they would be detected as being anomalous to the unknown natural clutter background. Using the specific algorithms discussed in this dissertation, all types of manmade objects, including the ones of interest to the users, and some natural clutter objects would be detected as anomalies in a natural background scene—as long as anomalous objects are present in the scene and are spectrally distinct from spectra of the spatially dominant clutter background. Another advantage of an effective anomaly detection algorithm is that potential targets yet unknown to the user could also be detected using such an algorithm.

The users and researchers could benefit from practical HS anomaly detection algorithms using the modeling strategies of this dissertation.

References

[1] J. B. Campbell. *Introduction to Remote Sensing,* 2nd ed. New York: Guilford, 1996.

[2] T. M. Lillesand and R. W. Kiefer. *Remote Sensing and Image Interpretation.* New York: Wiley, 1994.

[3] R. A. Schowengerdt. *Remote Sensing, Models and Methods for Image Processing,* 2nd ed. San Diego, CA: Academic Press, 1997.

[4] D. Manolakis, D. Marden, and G. Shaw, "Target detection algorithms for hyperspectral imaging applications," *Lincoln Laboratory Journal*, 14(1), 2003.

[5] Schott, J. R., "Remote Sensing, The Image Chain Approach", Oxford University Press, New York, NY, 1997

[6] Fukunaga, K., "Introduction to Statistical Pattern Recognition", 2nd edition, Academic Press, Inc., San Diego, CA, 1990.

[7] E. Crist, C. Schwartz, and A. Stocker, "Pairwise adaptive linear matched-filter algorithm," in *Proc DARPA Adaptive Spectral Reconnaissance Algorithm Workshop,* Jan 1999.

[8] Heesung Kwon, and Nasser M. Nasrabadi, "Kernel Spectral Matched Filter for Hyperspectral Imagery," J. of Computer Vision, Vol. 71, No. 2, pages 123-253, Feb. 2007.

[9] J. Grossmann, J. Bowles, D. Hass, J. Antoniades, M. Grunes, P. Palmadesso, D. Gillis, K. Tsang, M. Baumback, M. Daniel, J. Fisher, and I. Triandaf "Hyperspectral analysis and target detection system for the adaptive-spectral reconnaissance program (ASRP)," *Proc SPIE,* vol. 3372, pp.2-13, Apr. 1998.

[10]   C. Chang, X. Zhao, M. Althouse, and J. Pan, "Least squares subspace

projection approach to mixed pixel classification for hyperspectral images," *IEEE*

Trans. Geosci. Remote Sensing, vol. 36, pp. 898-912, May 1998.

[11]   D. Slater and G. Healey, "Exploiting an atmospheric model for automated

invariant material identification in hyperspectral imagery," *Proc. SPIE,* vol. 3372,

pp. 60-71, Apr. 1998.

[12]   A. Stocker, "Stochastic expectation maximization (SEM) algorithm," in *Proc.*

DARPA Adaptive Spectral Reconnaissance Algorithm Workshop, Jan. 1999.

[13]   P. Masson and W. Pieczynski, "SEM algorithm and unsupersivsed statistical

segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing,* vol. 31,

pp. 618-633, May 1993.

[14]   X. Yu, L. Hoff, I. Reed, A. Chen, L. Stotts, "Automatic target detection and

recognition in multiband imagery: A unified ML detection and estimation

approach," *IEEE Tran. Image Processing,* vol. 6, pp. 143-156, Jan. 1997.

[15]   H. Kwon, N. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector

for hyperspectral," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no.

2, February, 2005.

[16]   H. Kwon, S.Z. Der, and N.M. Nasrabadi, "Adaptive anomaly detection using

subspace separation for hyperspectral imgery," *Opt. Eng.,* v. 42 (11), November

2003, pp. 3342-3351.

[17]   B. Schokopf, A. J. Smola, and K-R Muller, "Kernel principal component

analysis," *Neural Computation*, (10):1299-1319, 1999.

[18]   M. Porter and H. Enmark, "A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)," *SPIE* vol. 834, Imaging Spectroscopy II, pp. 22-31, 1987.

[19]   G. Vane, R. O. Green, T. Go. Chrien, H. T. Enmark, E. G. Hansen and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of the Environment,* vol. 44, pp. 127-143, 1993.

[20]   Surface Optics Corporation, Website: http://www.surfaceoptics.com

[21]   J. R. Schott. *Remote Sensing: The Imaging Chaing Approach.* Oxford University Press, New York, 1997.

[22]   R.O. Duda, P.E. Hart. *Pattern Classification Scene Anal.*, Second Edition, New York: J. Wiley & Sons, 2004.

[23]   A. M. Law, W. D. Kelton. *Simulation Modeling and Analysis*, Third edition, Boston, MA: McGraw –Hill, 2000, pp. 363.

[24]   S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: clutter adaptation in anomaly detection," *IEEE Trans. Information Theory* August 2000, *46*, no. 5, pp. 1855–1871.

[25]   E. L. Lehmann. *Testing Statistical Hypotheses*, New York: transferred to Chapman & Hall, 2nd Edition, 1993.

[26]   A. Stocker, I. S. Reed, and X. Yu, "Multidimensional signal processing for electro-optical target detection," *Proc. SPIE,* Orlando, FL, vol. 1035, Apr 1990.

[27]   A. Margalit and I. S. Reed, "Adaptive detection of stationary optical and IR targets using correlated scene," Ph.D dissertation, Univ. of Southern Calif., Los Angeles, Nov. 1984.

[28]   V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[29]   G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, (12):2385-2404, 2000.

[30]   J. Qin and B. Zhang, "A goodness of fit test for logistic regression models based on case-control data," *Biometrika,* vol 84, 609-618, 1997.

[31]   K. Fokianos, B. Kedem, J. Qin, "A semiparametric approach to the one-way layout," *Technometrics*, pp. 56--65, 2001.

[32]   H. Scheffe. *The Analysis of Variance,* New York: John Wiley & Sons, 1959, Chapter 3.

[33]   D. Rosario and P. Rauss, "Experimental Results Detecting Camouflaged Soldiers in VNIR Hyperspectral Imagery," in *Proc. 2005 Military Sensing Symposia (MSS)*, 14-18 Feb 2005.

[34]   J. P. Egan. *Signal Detection Theory and ROC Analysis,* New York: Academic Press, 1975. J. P. Egan. *Signal Detection Theory and ROC Analysis,* New York: Academic Press, 1975.

[35]   The MATLAB MathWorks Website: http://www.mathworks.com/

[36]   K-N. Liou. *An Introduction to Atmospheric Radiation*, vol 26 of International Geophysics Series, San Diego, CA: Academic Press, Inc., 1980.